

Technical Report 1343

**Selecting Soldiers and Civilians into the U.S. Army
Officer Candidate School: Developing Empirical
Selection Composites**

Matthew T. Allen

Bethany H. Bynum

Rebecca T. Erk

Human Resources Research Organization

Nehama E. Babin

Mark C. Young

U.S. Army Research Institute

July 2014



**United States Army Research Institute
for the Behavioral and Social Sciences**

Approved for public release; distribution is unlimited

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

**Department of the Army
Deputy Chief of Staff, G1**

Authorized and approved for distribution:

**MICHELLE SAMS, Ph.D.
Director**

Research accomplished under contract
for the Department of the Army by:

Human Resources Research Organization

Technical review by:

Agata M. Gluszek, U.S. Army Research Institute
Irwin Jose, U.S. Army Research Institute

NOTICES

DISTRIBUTION: This Technical Report has been submitted to the Defense Information Technical Center (DTIC). Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-MS, 6000 6th Street, Bldg 1464 / Mail Stop 5610, Fort Belvoir, Virginia 22060-5610.

FINAL DISPOSITION: Destroy this Technical Report when it is no longer needed. Do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE					
1. REPORT DATE (dd-mm-yy) July 2014		2. REPORT TYPE Final		3. DATES COVERED (from. . . to) February 2011 to December 2012	
4. TITLE AND SUBTITLE Selecting Soldiers and Civilians into the U.S. Army Officer Candidate School : Developing Empirical Selection Composites				5a. CONTRACT OR GRANT NUMBER W91WAW-09-0013	
				5b. PROGRAM ELEMENT NUMBER 62785	
6. AUTHOR(S) Matthew T. Allen, Bethany H. Bynum, Rebecca T. Erk; Nehama E. Babin, Mark C. Young				5c. PROJECT NUMBER A790	
				5d. TASK NUMBER 329	
				5e. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Human Resources Research Organization 66 Canal Center Plaza, Suite 700 Alexandria, Virginia 22314				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Bldg 1464 / Mail Stop: 5610) Ft. Belvoir, VA 22060-5610				10. MONITOR ACRONYM ARI	
				11. MONITOR REPORT NUMBER Technical Report 1343	
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution Statement A: Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES Contracting Officer's Representative and Subject Matter Expert POC: Dr. Nehama E. Babin					
14. ABSTRACT (<i>Maximum 200 words</i>): The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) initiated the <i>Measures for Selecting Soldiers for the Officer Candidate School</i> program of research to help select OCS applicants with high leadership potential and likelihood of staying in the Army. The present effort reports results from the third phase of this research. The purpose of this phase was to (a) cross-validate the results from the second phase with a sample of newly accessioned OCS candidates, (b) select the most promising individual instrument for predicting officer performance and continuance, and (c) develop an empirical selection composite. Results suggest that the Phase 2 and Phase 3 samples were quite different on a number of demographic and performance characteristics. Despite these differences, results suggest that the predictive validity for the experimental measures was consistent between Phase 2 and Phase 3. The most promising test measured personality and motivations. Finally, a new set of composites developed to predict officer performance and continuance demonstrated strong validity that held up in a different sample from the development sample. ARI plans to use these results to prepare the experimental measures for testing in an operational setting.					
15. SUBJECT TERMS Manpower , Selection and classification, Officer Candidate School, Officer selection, Longitudinal analysis					
SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT Unlimited Unclassified	20. NUMBER OF PAGES 102	21. RESPONSIBLE PERSON Dorothy Young (703) 545-2316
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified			

Technical Report 1326

**Selecting Soldiers and Civilians into the U.S. Army
Officer Candidate School: Developing
Empirical Selection Composites**

**Matthew T. Allen
Bethany H. Bynum
Rebecca T. Erk**

Human Resources Research Organization

**Nehama E. Babin
Mark C. Young**
U.S. Army Research Institute

**Personnel Assessment Research Unit
Tonia S. Heffner, Chief**

**U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street, Building 1464
Fort Belvoir, Virginia 22060**

July 2014

Approved for public release; distribution is unlimited.

ACKNOWLEDGEMENTS

The authors would like to thank the numerous individuals who contributed to the successful completion of this project. Ani DiFazio and Christina Thibodeaux of HumRRO, Sharon Meyers, Veronica Gilrane, and Lindsey Bartholomew of ARI, and Independent Contractor Winnie Young were all instrumental in collecting, encoding, cleaning, and finalizing the data for analysis.

At Fort Benning, Barbi Carroll was instrumental in coordinating data collections and other activities with OCS personnel. The company commanders whose candidates participated in this research were also very helpful in this effort. The authors are also indebted to LTC Dean Bushnell and LTC John Best, the two OCS Commandants whose tenure overlapped with this research, for their support in this effort.

Finally, the authors would like to thank the OCS Officer Candidates who agreed to participate in this research. Without their support, this research would not be possible.

SELECTING SOLDIERS AND CIVILIANS INTO THE U.S. ARMY OFFICER CANDIDATE SCHOOL: DEVELOPING EMPIRICAL SELECTION COMPOSITES

EXECUTIVE SUMMARY

Research Requirement:

The U.S. Army Officer Candidate School (OCS) needs candidates who are likely to stay in the Army and perform well as leaders. With that in mind, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) initiated a research program in 2008 called *Measures for Selecting Soldiers for the Officer Candidate School* (referred to as “SelectOCS”). The overarching objective of SelectOCS is to identify a test battery likely to select OCS applicants with promising leadership potential, strong fit with the Army, and high likelihood of staying in the Army. The subject of this report is the activities that comprised the third phase of the SelectOCS program of research.

SelectOCS Phase 1 involved developing a battery of measures called the Officer Background and Experiences Form (OBEF) and collecting data from 1,344 officer candidates from 10 OCS classes (Russell & Tremble, 2011). Validation results pointed to a number of OBEF measures that could improve the prediction of candidate commitment, career intentions, and even OCS performance beyond the level of prediction afforded by the Armed Forces Qualification Test (AFQT). There were some differences in the relative strength of the individual predictor measures depending upon whether the candidates came to OCS through (a) the *enlistment option* program, in which civilians with college degrees enter OCS after completing Basic Combat Training (BCT); or (b) the *in-service* program, in which enlisted Soldiers are selected for OCS participation. Note that the in-service program is the traditional route to OCS, while the Army launched the enlistment option program in 1998. Separate predictor composites comprised of scales from the OBEF were developed for the in-service and enlistment option candidates. Analyses indicated that the composites predicted the targeted OCS criteria at generally comparable levels for the two groups. Subsequent analyses indicated that these composites also predict officer performance and continuance 3 years after completing OCS (Allen & Young, 2012).

The follow-up work conducted in SelectOCS Phase 2 expanded on the Phase 1 results in a number of ways (Russell, Allen, & Babin, 2011). First, a new OBEF was constructed and administered to 807 candidates in five OCS classes. The OBEF built on measures demonstrating promise in Phase 1 and other ARI-sponsored efforts to predict Army officer performance and continuance. As with Phase 1, these OBEF measures were validated against a number of criterion measures, including Peer Ratings of Leadership Potential and self-reported active duty career intentions. Results of the Phase 2 analyses suggest that the predictor composites developed as part of Phase 1 demonstrated reasonable levels of predictive validity in Phase 2. However, a number of the new scales also demonstrated promise for predicting key criteria of interest, suggesting the composites could be further revised to enhance predictive validity.

With SelectOCS Phase 1 and Phase 2 results in mind, the purpose of the current Phase 3 effort was to (a) cross-validate the results from SelectOCS Phase 2 with a sample of newly accessioned OCS candidates, (b) select the most promising individual OBEF instrument for

predicting officer performance and continuance, and (c) develop a revised set of empirical selection composites.

Procedure:

The procedures in the current research (Phase 3) mirror those of SelectOCS Phase 2 (Russell et al., 2011). The OBEF comprised five assessments: (a) a variant of the Rational Biodata Inventory (RBI) – a biographical instrument measuring personality, commitment, and motivation; (b) the Tailored Adaptive Personality Assessment System (TAPAS) – a forced-choice personality assessment; (c) a Work Values measure comprised of rank-ordered and Likert-scale items; (d) the Army Identity Structure scales – graphical items designed to measure deep structure of an individual's identification with the Army; and (e) the Leader Knowledge Test (LKT) – a measure of implicit leadership. In Phase 3, five scales were added to the RBI, and one scale was dropped from the work values instrument. The OBEF was administered to all candidates in five classes at the end of their first week at OCS ($n = 459$). We then collected criterion data through an end-of-class (EOC) survey and from OCS administrative records. The EOC survey included measures of (a) Affective Commitment to the Army, (b) Active Duty Career Intentions, and (c) Leadership Potential as rated by candidates' peers. Candidates' Armed Services Vocational Aptitude Battery (ASVAB) scores were obtained from Army personnel records. All of the predictor and criterion measures used in this research exhibited acceptable psychometric properties.

Findings:

Objective 1: Cross-Validate Phase 2 Results

The Phase 2 and Phase 3 samples differed in terms of demographic variables, OBEF scores, and performance outcomes. The Phase 3 sample had a much higher percentage of enlistment option candidates than did Phase 2 (72% versus 56%), more advanced degrees, and higher AFQT scores. Phase 3 had slightly more white, non-Hispanic males, and tended to be younger. In terms of OBEF scores, where differences existed, the Phase 3 sample tended to have significantly higher scores on positively valenced scales and lower scores on negatively valenced scales as compared to the Phase 2 sample. The Phase 3 sample also tended to have higher scores on the key performance and continuance criteria, such as Affective Commitment and Peer Ratings of Leadership Potential. Despite the differences between the Phase 2 and Phase 3 samples, results of the equivalence analyses suggest that, for all of the OBEF instruments, the predictive validity results were very consistent between the two samples, with the exception of one relationship. There was a significant difference between Phase 2 and 3 in how the RBI predicted candidate physical fitness, as measured by their Army Physical Fitness Test (APFT) scores, such that the magnitude of the correlation coefficient was higher in Phase 3. However, the R^2 value between the RBI and APFT were statistically significant and large for both samples, suggesting that the predictive relationship was stable over time. Given the majority of the relationships were similar for the Phase 2 and Phase 3 samples, we combined the two samples for the incremental validity and composite formation analyses.

Objective 2: Select the Most Promising OBEF Instrument

Results based on the incremental validity of each instrument beyond the AFQT for predicting target criteria suggest that the RBI is the most promising single measure for predicting key outcomes in both the in-service and enlistment option samples. The next most promising measure overall was the TAPAS, followed by the Work Values, and Army Identity Structure scales. Based on these results, the RBI formed the basis for developing a set of empirical selection composites.

Objective 3: Develop Empirical Composites

The goal for developing empirical selection composites was to maximize prediction while minimizing the number of scales included in the composite. In general, the OBEF scales that predicted continuance criteria (Affective Commitment, Career Intentions) differed from the OBEF scales that predicted performance criteria (Academic, Leadership, Fitness). Therefore, we developed separate composites for each outcome category. Due to small sample size in the in-service sample ($n = 36$ in the Phase 3 sample) and their diminishing role in OCS,¹ the composites were developed only for candidates who entered OCS through the enlistment option (Phase 2 $n = 429$; Phase 3 $n = 319$). Formation of each composite involved the following steps: (a) developing weighted RBI composites, (b) identifying scales to add to the RBI composites, (c) constructing a final composite, and (d) evaluating that composite.

The final composites included a diverse array of OBEF predictors.² The new composites accounted for an additional 7.8% to 31.4% of the variance beyond that afforded by the AFQT on all key outcomes of interest except Academic performance. More importantly, validity decreased little in a separate holdout sample, suggesting these composites are stable predictors of key outcomes. The exception to this was the prediction of Career Intentions, where the ΔR decreased from .33 in the analysis sample to .09 in the holdout sample. However, this was primarily due to an increase in the predictive efficacy of the AFQT in the holdout sample.

Utilization and Dissemination of Findings:

The results presented here suggest that the OBEF holds considerable promise for identifying applicants with high and low potential for performance and for staying in the Army beyond their initial Active Duty Service Obligation (ADSO). It also demonstrates which components of the OBEF have the highest potential (in terms of predictive validity) to add value to the Army's current system for selecting candidates into OCS.

In the future, the final composites could be used in several different ways. A multiple-hurdle approach would be to set a cut score on one composite (e.g., perhaps a very low score to deselect the poorest candidates on it) then to select candidates on the other composites from the remaining pool. Each composite could have its own cut score or composite scores could be

¹ In FY 2011, the goal was to commission about 1,722 lieutenants from OCS. Most (1,200) of the FY 2011 goal were to come from the enlistment option program while 522 were targeted from the in-service program (MILPER 10-164, 23 June 2010, <https://www.benning.army.mil/199th/ocs/index.htm>)

² Due to their sensitive nature, we do not provide the specific OBEF scales included in each composite in this report. Those interested in obtaining a copy of this information should contact the report authors.

combined. Another approach would be to weight the composite scores based on the Army's priorities and form an overall weighted composite. These options could be explored using data from an operational try-out. ARI plans to try out selected OBEF scales in an operational sample of applicant for OCS. Future research will also follow-up on the Phase 1, 2, and 3 samples to explore whether the OBEF also predicts long-term officer outcomes, such as in-unit performance and actual continuance.

SELECTING SOLDIERS AND CIVILIANS INTO THE U.S. ARMY OFFICER CANDIDATE SCHOOL: DEVELOPING EMPIRICAL SELECTION COMPOSITES

CONTENTS

	Page
Chapter 1: Background and Overview.....	1
Officer Candidate School Overview	1
The SelectOCS Research Program.....	3
SelectOCS Phase 1.....	4
SelectOCS Phase 2.....	5
SelectOCS Phase 3 (Current Effort)	5
Predictors of Officer Performance and Continuance	6
Proximal Predictors.....	7
Organization of Report.....	8
Chapter 2: Research Method.....	9
Overview	9
Design and Procedures	9
Database Development.....	10
Participants	10
Predictor Instruments	13
Armed Forces Qualification Test (AFQT).....	13
Rational Biodata Inventory (RBI).....	13
Tailored Adaptive Personality Assessment System (TAPAS)	14
Work Values	16
Leader Knowledge Test (LKT).....	17
Army Identity Structure	18
Criterion Measures	18
End-of-Class (EOC) Measure	18
OCS Performance Data.....	20
Subgroup Differences on Predictors.....	22
Summary	23
Chapter 3: Predicting Candidate Performance with the OBEF	24
Analytic Approach	24
Missing Data Estimation.....	25
Differential Prediction	25
Incremental Validity	26
Range Restriction.....	27
Shrinkage	28

CONTENTS (CONTINUED)

	Page
Results	28
Cross-validation of Phase 2 and Phase 3 Results.....	28
Incremental Validity	33
Summary and Conclusions	38
Chapter 4: Development and Evaluation of Revised OBEF Composites	40
Approach	40
Procedure	40
Results	42
RBI Composites	42
Full Composite.....	45
Cross-Validation	48
Subgroup Differences	48
Conclusion.....	49
Chapter 5: Conclusions and Next Steps.....	55
Background and Objectives.....	55
Summary of Results	55
Objective 1: Cross-Validate Phase 2 Results.....	56
Objective 2: Select the Most Promising OBEF Instrument	56
Objective 3: Develop Empirical Composites.....	56
Implications	57
For Practice	57
For Research	59
Limitations.....	60
Future Directions	60
References	62

Appendices

Appendix A - Manual for Collecting Data in CY2011 for SelectOCS Phase 3	A-1
Appendix B - Descriptive Statistics and Subgroup Differences	B-1
Appendix C - Bivariate Predictor and Criterion Intercorrelations for the SelectOCS Phase 2 and 3 Samples	C-1
Appendix D - Full Information Maximum Likelihood (FIML) Bivariate Correlations for Phase 3 In-Service and Enlistment Option Candidates.....	D-1

Tables

Table 2.1. Officer Candidate School (OCS) Beginning-of-Class (BOC) Sample Sizes and End-of-Class (EOC) Response Rates for the SelectOCS Phase 2 and 3 Samples	9
Table 2.2. Officer Candidate School (OCS) Demographics for SelectOCS Phase 2 and 3 Samples	11
Table 2.3. Officer Candidate School (OCS) Continuous Demographics for SelectOCS Phase 2 and 3 Samples	12
Table 2.4. RBI Scales Administered in Phase 3	15
Table 2.5. TAPAS Scales by Big Five Trait	16
Table 2.6. Phase 3 Work Values Content	17
Table 2.7. End-of-Class Criterion Variables	19
Table 2.8. OCS Administrative Criterion Variables	21
Table 3.1. Cohort Differences between the Phase 2 and Phase 3 Enlistment Option Cohorts on OBEF Scales	29
Table 3.2. Measurement Invariance Cross-Validation Analyses of Phase 2 and Phase 3 Enlistment-Option Cohort	32
Table 3.3. Incremental Validity Results for the Enlistment Option Sample with Phase 2 and Phase 3 Samples Combined	34
Table 3.4. Incremental Validity Results for the In-Service Sample with Phase 2 and Phase 3 Samples Combined	37
Table 4.1. Relative Importance of the RBI Scales for Predicting Select Continuance Criteria in the Analysis Sample	43
Table 4.2. Relative Importance of the RBI Scales for Predicting Select Performance Criteria in the Analysis Sample	44
Table 4.3. Incremental Validity Results for the RBI Composites	45
Table 4.4. Bivariate Correlations, Semi-Partial Correlations, and Relative Weights for the Continuance Composite	46
Table 4.5. Bivariate Correlations, Semi-Partial Correlations, and Relative Weights for the Performance Composite	47
Table 4.6. Incremental Validity Results for the Full Continuance Composite	48
Table 4.7. Incremental Validity Results for the Full Performance Composite	49
Table 4.8. Subgroup Differences on Continuance and Performance Composites for the Enlistment Option Cohort	50

SELECTING SOLDIERS AND CIVILIANS INTO THE U.S. ARMY OFFICER CANDIDATE SCHOOL: DEVELOPING EMPIRICAL SELECTION COMPOSITES

Chapter 1: Background and Overview

The U.S. Army places thousands of young men and women in positions of leadership. These positions often entail responsibility for supervising hundreds of Soldiers in highly complex and dangerous environments. The Army is unique from other organizations, both in the scope of responsibility afforded to young officers, and in the number of officers that must be trained to effectively operate in positions of leadership. It is not surprising then that the officer personnel system in the Army must be quite large and complex to maintain the size and quality of its officer corps. According to the Department of Defense, in late 2011 there were approximately 81,000 commissioned officers in the 560,000+ Regular Army.³ Given these force requirements, the Army's challenge is to select, develop, and retain officers who can meet current requirements and adapt to future missions.

An individual can become a commissioned Army officer through four avenues: The Reserve Officer Training Corps (ROTC), the U.S. Military Academy at West Point (USMA), Officer Candidate School (OCS), and direct commission. The first three commission individuals as officers upon graduation, when they enter the Army as second lieutenants (pay grade O-1). A relatively small number of individuals enter the officer corps through the "direct commissioning" process. These individuals typically have highly developed skills prior to entry (e.g., medical, legal, chaplain) and are assigned to special branches. Through these commissioning sources (particularly ROTC, USMA, and OCS), the Army invests considerable resources to prepare officers to take on these positions of leadership, and thus is particularly interested in selecting officers who can internalize the rigorous training required, can perform well in positions of leadership in their units, and will consider remaining in the Army beyond their service obligation.

For many years, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has been conducting research to help the Army select officers from a pool of applicants that meet these requirements. The current effort seeks to contribute to this corpus of research by developing empirical composites using a battery of measures developed in previous research, for selecting candidates into OCS. We begin this chapter with a description of OCS and its current selection and accessioning process, followed by a brief review of the extant literature on predicting leadership performance in a military setting. We conclude with an overview of the remaining chapters in this report.

Officer Candidate School Overview

OCS is a 12-week course with academic (e.g., military history courses), physical fitness (e.g., runs with and without gear, obstacle course runs), and leadership (e.g., rotational assignments to leadership positions) components. Historically, the Army has used OCS to fill

³ This number does not include warrant officers, see <http://siadapp.dmdc.osd.mil/personnel/MILITARY/rg1111.pdf>. Retrieved 1/27/2012.

accession gaps in officer requirements when other commissioning sources cannot meet force structure requirements. Compared to ROTC and the USMA, OCS is flexible enough to increase or decrease its production of officers on short notice. Particularly in wartime, the Army must make the most of all officer-accessioning sources to attract capable officers with strong propensities for retention. Between 1998 and 2008, OCS expanded significantly. In 2006-2008 OCS accounted for approximately 35% to 40% of new Army officer accessions, whereas historically only about 10% to 15% of officer accessions came from OCS (Henning, 2006; Wardynski, Lyle, & Colarusso, 2009). During this time, OCS grew through the increased use of an alternative avenue for entry into OCS: the “enlistment option” program. This option involves recruiting civilians who have a college degree to enter OCS immediately after joining the Regular Army as enlisted Soldiers and completing Basic Combat Training (BCT). The Army refers to these candidates as “enlistment” option because they are technically enlisted Soldiers for a short time period. This program supplements the traditional “in-service” route to OCS in which enlisted Soldiers (typically non-commissioned officers [NCOs]) apply for OCS participation. By 2009, the enlistment option accounted for about half of all OCS commissions (Wardynski et al., 2009), and it accounted for more than 70% of the sample collected in 2011 as part of the current effort (see Chapter 2).

Despite its current importance to maintaining Army force requirements, researchers have dedicated little effort to the selection of candidates into OCS in the past 30 years. However, researchers performed substantial work developing and validating selection tools for use in OCS prior to 1979 (Rumsey, 2012). According to Rumsey (2012), the greatest emphasis on OCS selection research occurred between 1941 (at the beginning of U.S. involvement in World War II) and 1957. Consistent with personnel research conducted at the time, the early selection instruments measured cognitively oriented aspects of officer candidates, such as general cognitive aptitude, data interpretation, arithmetic reasoning, and reading comprehension. Later assessments used alternative approaches (e.g., biographic self-reports and structured interviews) to predict candidate leadership performance and career continuance. This research ultimately led to the development of the Officer Selection Battery (OSB),⁴ which was administered to incoming OCS candidates in 1975. The Army officially authorized the use of the OSB for selecting OCS candidates in 1979, where it remained for a number of years. However, no other selection tool replaced the OSB for screening OCS candidates after it was dropped from use.

The current eligibility requirements for both in-service and enlistment option applicants for OCS include (a) an Armed Services Vocational Aptitude Battery (ASVAB) General Technical (GT) score of 110 or above, (b) a passing score on the Army Physical Fitness Test (APFT), and (c) a 4-year Baccalaureate degree from an accredited college or university (Oliver, Ardison, Russell, & Babin, 2011). Although measures of cognitive aptitude such as the ASVAB have proven to be excellent predictors of technical proficiency (e.g., performance on a work sample test), previous research suggests that it is less predictive of non-technical aspects of Soldier performance, such as effort, leadership, and personal discipline (Campbell & Knapp, 2001; Knapp & Heffner, 2010). Other than these eligibility requirements, OCS candidate selection relies primarily on qualitative and descriptive materials such as interviews,

⁴ Earlier versions of the OSB included the Cadet Evaluation Battery (CEB) and the Officer Leadership Qualification Inventory (OLQ).

documentation of educational requirements, recommendations from superiors, and essays asking individuals why they want to become Army officers. The qualitative assessments are in place to assess whether the individual applying to OCS has the character to be an effective Army officer. However, there is little standardization in how the Army administers and scores these instruments, with some judges being more lenient and others more severe.

In summary, we can draw the following inferences about the OCS selection and accessioning process:

1. OCS is a significant source for new officers in the U.S. Army and, for at least the immediate future, plays a prominent role in fulfilling force requirements.⁵
2. OCS has successfully administered instruments such as the OSB that predict officer performance and continuance, and has used them to make operational decisions. However, these instruments are no longer part of the current OCS selection process.
3. Current methods for selecting OCS candidates do not allow for systematically (a) screening out individuals with little performance or continuance potential or (b) identifying high potential individuals.
4. No previous research has examined the characteristics of effective officers entering through the newer (and currently more prominent) enlistment option program.

Recognizing this lack of current research into the OCS selection process and the potential value additional assessment methods could have on helping OCS meet its mission, ARI initiated a new research program termed “SelectOCS.”

The SelectOCS Research Program

The research described in this report is part of a broader, multi-year effort by ARI to identify, develop, and validate personnel tests for use in officer selection. The current project is the third phase of the *Measures for Selecting Soldiers for the Officer Candidate School* (referred to in the remainder of this report as SelectOCS) research effort, which focused on officers entering via OCS. A parallel research effort focused on officers entering the Army via ROTC (e.g., Putka, 2009; Putka, Kilcullen, Tremble, Wasko, & Shaw, 2009). The primary goal of both the SelectOCS and ROTC efforts has been to identify, recommend and, eventually, implement selection measures that will increase the likelihood of accessioning officers who perform well at both junior and senior grades and intend to make the Army a career.

The SelectOCS research program began in 2008 with the development of the first Officer Background and Experiences Form (OBEF), a battery of measures designed to assess aspects of applicant personality, values, and judgment. These mostly “non-cognitive” domains stand in contrast to the Army’s primary instrument for selection – the ASVAB – a measure of applicant knowledge and aptitude in domains such as mathematics. Given the ASVAB is already

⁵ Note that as of this writing, OCS was in the process of reducing the number and size of its classes.

embedded in the Army's current selection and accessioning processes, it is critical that the OBEF measure aspects of individuals not already measured by the ASVAB. The current effort is the third phase in a research program designed to validate the OBEF for the purposes of selecting candidates into OCS.

SelectOCS Phase 1

The specific objectives of SelectOCS Phase 1 were to (a) develop and validate a predictor battery for identifying OCS applicants with the most leadership potential, the best fit with the Army, and the greatest likelihood of staying in the Army beyond their initial Active Duty Service Obligation (ADSO); and (b) investigate the outcomes of two accession options (i.e., in-service and enlistment option) to OCS (Russell & Tremble, 2011). Toward that end, ARI, in collaboration with the Human Resources Research Organization (HumRRO), designed the OBEF and administered it to 1,344 OCS candidates in 10 classes. The OBEF included measures that had demonstrated promise in previous efforts, as well as new experimental measures. The measures included (a) a variant of the Rational Biodata Inventory (RBI), a biographical instrument measuring personality; (b) a Work Values instrument representing work preferences investigated in prior officer and enlisted selection research; and (c) experimental measures of other non-cognitive attributes such as organizational identity, affectivity, and leadership judgment.

Researchers validated the OBEF against several criterion measures obtained through self-reports collected at the end of OCS training or through extractions of archival performance data from administrative files. Although a number of such measures were assembled, the criterion measures used throughout the effort consisted of either OCS class performance scores (academic, leadership, fitness, and total performance scores) or end-of-course self-reports of commitment to the Army and intentions for a regular Army career. Researchers collected validation data in Phase 1 from 609 candidates in six classes who graduated from the 12-week OCS course and for whom criterion data was obtainable at the time of report preparation.

Validity results pointed to a number of OBEF measures that could provide improvement to the prediction of candidate commitment, career intentions, and OCS performance beyond the level of prediction afforded by the Armed Forces Qualification Test (AFQT), a composite of multiple ASVAB subtests. Two sets of OBEF measures consistently stood out: (a) the RBI scales and (b) the work values scales. There were some differences between the in-service and enlistment option candidates in the relative strength of the individual predictor measures. These differences guided development of separate (though similar in content) composites for the in-service and enlistment option candidates. Analyses indicated that the composites (a) exhibited acceptable statistical properties (e.g., high reliability and few subgroup differences) and (b) predicted the targeted OCS criteria at generally comparable levels for the two groups.

ARI is currently conducting a follow-up investigation of the SelectOCS Phase 1 sample to determine whether the OBEF predicts officer performance and continuance beyond the training environment (Allen & Young, 2012). Preliminary results suggest that the RBI and work values components of the OBEF remained strong predictors of both in-unit officer performance (as determined by self-report assessments) and continuance beyond their initial ADSO. Furthermore, the in-service and enlistment option composites developed in Phase 1 also

predicted key outcomes of interest, such as technical performance, physical fitness, and commitment to the Army, suggesting the constructs included in these composites are stable predictors of officer performance over time.

SelectOCS Phase 2

The follow-up work conducted in Phase 2 expanded on the Phase 1 results in a number of ways (Russell, Allen, & Babin, 2011). First, a new OBEF was constructed and administered to 807 candidates in five OCS classes. The OBEF built on measures demonstrating promise in Phase 1 and other ARI-sponsored efforts to predict Army officer performance and continuance (e.g., Putka, 2009). Specifically, (a) new leadership motivation scales were added to the RBI to address content gaps in predicting leadership performance and potential; (b) two promising measures – a rank-ordered work values measure and the Leader Knowledge Test (LKT) – from other officer selection work were added (e.g., Allen, Thibodeaux, & Babin, 2010; Putka et al., 2009); and (c) a version of a measure that has shown promise for selecting enlisted Soldiers – the Tailored Adaptive Personality Assessment System (TAPAS) – was added to determine whether results held up in an officer sample (Knapp & Heffner, 2009; 2010). As with Phase 1, researchers validated the OBEF measures against a number of criterion measures, including Peer Ratings of Leadership Potential and current self-reported active duty career intentions.

Results of the Phase 2 analyses suggested that (a) the RBI and TAPAS generally demonstrated comparable levels of predictive validity, both with and without controlling for AFQT in the model, though the RBI did provide more incremental validity beyond the TAPAS than the TAPAS did beyond the RBI; (b) the predictor composites developed as part of Phase 1 demonstrated reasonable levels of predictive validity in Phase 2; and (c) a number of the new scales demonstrated promise for predicting key criteria of interest, suggesting the need to develop revised OBEF-based predictor composites (Russell et al., 2011).

SelectOCS Phase 3 (Current Effort)

The present effort sought to accomplish the following objectives:

1. Cross-validate the results from Phase 2 with a sample of newly accessioned OCS candidates. Cross-validation ensures that the results from Phase 2 can be explained by the predictive efficacy of the OBEF rather than an experimental or statistical artifact.
2. Select the most promising individual instrument (e.g., RBI, TAPAS, LKT) for predicting officer performance and continuance. Given the administration time restrictions in an operational setting, it is necessary to select the one measure in the OBEF that demonstrates the most promise for predicting key outcomes.
3. Develop an empirical selection composite or set of composites that balance both prediction of key criteria and parsimony. The composite(s) will use the measure selected in objective 2 above as the base, then use other OBEF scales to determine which measures could best add value to the base instrument. Again, parsimony (i.e., having fewer scales contributing to the composites) is also critical.

We accomplished these objectives by collecting new predictor and criterion data from five OCS classes. To prepare for the composite development activities, we briefly review the rationale behind the content of the selection battery. We then turn to describing the present effort in more detail.

Predictors of Officer Performance and Continuance

The terms “performance” and “continuance” are broad criteria for operationalizing whether an Army officer is successful in their role. “Performance” is a multidimensional term that denotes the behaviors that individuals exhibit on the job (Campbell, in preparation; Campbell, McCloy, Oppler, & Sager, 1993). In the Army context, previous treatments of officer performance have tended to emphasize leadership dimensions, but have also included other general and Army-specific dimensions, such as physical fitness and effort. “Continuance” is most analogous to “turnover” in the civilian sector. Unlike the civilian sector however, OCS-commissioned officers are obligated to complete an 8-year military service obligation (MSO), with a 3-year ADSO.⁶ Traditionally, very few officers separate before the end of their ADSO (e.g., Allen & Young, 2012). A successful selection battery will maximize the Army’s capability to select Soldiers who both perform well as leaders and are likely stay in the Army beyond their ADSO.

Models of leadership and turnover typically distinguish between proximal and distal antecedents of these outcomes (e.g., Hom, 2011; Van Iddekinge, Ferris, & Heffner, 2009; see Figure 1). For example, according to Van Iddekinge and colleagues (2009), individual differences such as cognitive ability and personality (distal antecedents) feed into mediators such as experiences in a leadership role (proximal antecedents), which in turn predict leadership performance. In the context of selecting officers with the highest potential, we are typically most interested in distal antecedents, such as personality and cognitive ability. However, to the extent possible, more proximal antecedents can be used either as part of a selection battery (e.g., motivation to lead; Chan & Drasgow, 2001) or as proxy criteria when the actual criteria of interest are not available (e.g., intentions to quit as a proxy for actual continuance; Hom, 2011). We briefly describe the distal and proximal antecedents that seem most promising for selecting officers with high leadership and continuance potential into OCS below.

⁶ See Army Regulations 350-100 and 135-91.

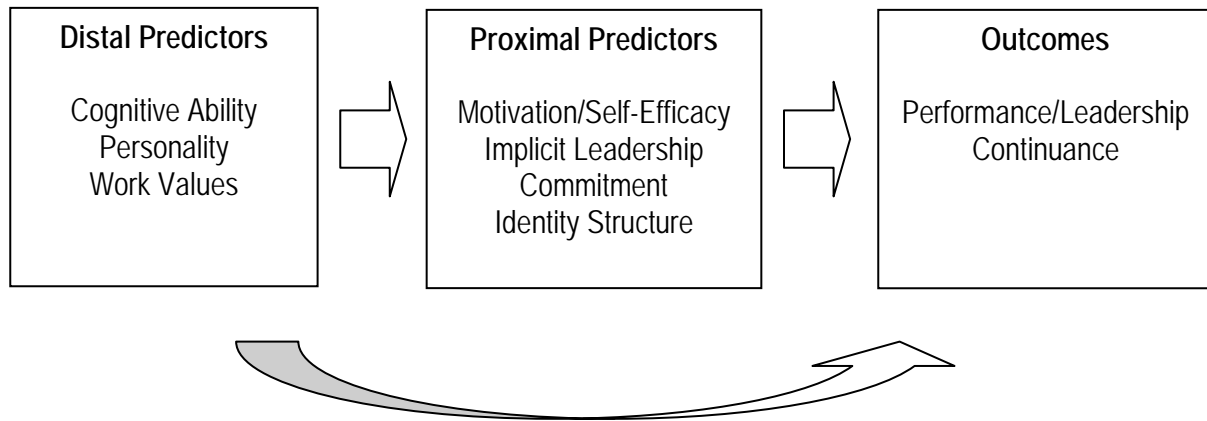


Figure 1. Conceptual model of predictor and criterion relations.

Distal Predictors

The present research includes three categories of individual difference instruments that we consider “distal” predictors of officer performance and continuance—cognitive ability, personality, and work values. General cognitive ability, or *g*, is a consistent predictor of training performance and performance on the job (Ree, Carretta, & Teachout, 1995; Schmidt & Hunter, 1998), particularly when researchers define “performance” by technical tasks, such as work sample tests (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Oppler, McCloy, & Campbell, 2001). Models of leadership performance generally hypothesize that the relations between cognitive abilities and leadership are mediated by knowledge and skill acquisition, a notion that has received some support in the literature (Connelly et al., 2000; Van Iddekinge et al., 2009). In contrast to performance, research has generally shown cognitive ability to be unrelated to turnover in the civilian sector (cf. Hom, 2011, Table 1). In the present effort, we assessed candidates’ general cognitive ability using AFQT scores gleaned from administrative records.

As with general cognitive ability, models of leadership performance tend to treat personality as a distal antecedent of key outcomes (Chan & Drasgow, 2001; Connelly et al., 2000; Van Iddekinge et al., 2009). For example, Chan and Drasgow (2001) found that personality predicted more proximal antecedents of leadership, such as leadership experience, leadership self-efficacy, and the motivation to lead. Additional meta-analytic studies of personality and leadership have found aspects of personality to be predictive of leadership performance (Judge, Bono, Ilies, & Gerhardt, 2002). With regard to turnover, previous research suggests that personality is a consistent predictor of more proximal antecedents of turnover such as job satisfaction and affective commitment (Allen & Young, 2012; Judge, Heller, & Mount, 2002; Schleicher, Hansen, & Fox, 2011). In the present effort, personality was assessed by two measures embedded in the OBEF—the TAPAS and the RBI.

A third distal antecedent of performance and continuance is an individual’s work values. The relationship between specific work values, such as valuing work that contributes to society and key outcomes depends on the nature of the job. For example, those who place little value in doing work that contributes to society would have values considered incongruent with the

Army's environment. Thus, these individuals are more likely to be dissatisfied, disengaged, and lower performing than individuals who place higher value on work that contributes to society. When conceptualized in this "person-environment" fit context, work values predict multiple outcomes such as job satisfaction and intentions to quit (see Schleicher, Hansen, & Fox, 2011). In the present effort, we assessed individual work values using the OBEF Work Values instrument.

Proximal Predictors

Candidate motivations and self-efficacy are generally thought to be more proximal predictors of key outcomes such as performance and continuance. For example, self-efficacy, or an individual's confidence in achieving a successful outcome, is a strong and consistent predictor of performance in a training setting (Colquitt, LePine, & Noe, 2000; Campbell & Kuncel, 2002). Additional studies expanded on this research by demonstrating that leadership self-efficacy (i.e., confidence that one can perform effectively in a leadership role) predicts another proximal antecedent of leadership performance—motivation to lead (Chan & Drasgow, 2001; Hendricks & Payne, 2007). Motivation to perform physical activities is also a consistently strong predictor of Soldier and officer performance (e.g., Knapp & Heffner, 2009; 2011; Putka, 2009; Putka et al., 2009; Wasko et al., 2011). We measured motivation and self-efficacy with the RBI.

Implicit leadership theories, or the underlying beliefs that an individual holds about what constitutes effective leadership, are related indirectly to improved relationships with subordinates and effective outcomes (Epitropaki & Martin, 2005). Those who have implicit theories that are more in line with the prototype Army officer are more likely to exhibit effective behaviors than those with implicit theories that are less in line with the prototype Army officer. To assess implicit leadership, the present project administered the LKT, which we also administered in SelectOCS Phase 2 and in the ROTC projects (Russell et al., 2011; Wasko et al., 2011). In contrast with implicit leadership, organizational commitment is a more proximal predictor of continuance rather than performance (Hom, 2011; Tett & Meyer, 1993; Strickland, 2005). Commitment also relates to individual employee performance, but the magnitude of the effect is generally weaker than for continuance (Karrasch, 2003). A distinct but related idea to commitment is the notion of "identity structure" (Robbins, Allen, & Putka, 2011), which is based on social identity theory (Tajfel & Turner, 1979, 1986). Although there is some overlap between measures of identity and affective commitment, identity is generally thought to comprise different components (e.g., identity stability) that might also be related to key continuance outcomes (Meyer, Becker, & Van Dick, 2006). We used the RBI to assess affective commitment to the Army, while we used separate "Army Identity Structure" items to assess identity structure.

Organization of Report

Recall that the objectives for the present effort are to (a) cross-validate the results from Phase 2 with a new sample of newly accessioned candidates, (b) select the most appropriate individual instrument for predicting officer performance and continuance, and (c) develop an empirical selection composite. Chapter 2 describes our method for collecting and processing the data to accomplish these tasks. We report the analyses conducted to address the first two objectives in Chapter 3, and we report the analyses conducted to accomplish the third objective in Chapter 4. In Chapter 5, we summarize these results in relation to these three key objectives. We also discuss implications and directions for future research.

Chapter 2: Research Method

Overview

The research method for SelectOCS Phase 3 mirrors that of the Phase 2 procedures described in Russell, Allen, and Babin (2011). We reiterate the methodology here, with an emphasis on the similarities and differences between the Phase 2 and Phase 3 samples and instruments.

Design and Procedures

The present effort used a longitudinal validation design, tracking candidates from the beginning of OCS to the end, with the intention of tracking them through the end of their ADSO. We administered the OBEF and a demographics form to officer candidates at the end of their first week of OCS with paper-and-pencil instruments proctored by two to four HumRRO and/or ARI staff members. To ensure the data collections were standardized, we created a data collection manual, which is included in Appendix A. We sent a solicitation to officer candidates through their Army Knowledge Online (AKO) email addresses to complete an online end-of-class (EOC) measure two weeks before the end of the 12-week course. We sent reminders to non-respondents in classes that had low initial response rates. We obtained informed consent for both the beginning-of-class (BOC) and EOC data collections.

The response rates for Phases 2 and 3 are summarized in Table 2.1. The table shows that both the overall sample sizes and the EOC response rates decreased in Phase 3. Note that in addition to non-responding, other factors (e.g., recycles during the course, early separations) can also adversely affect the response rates.

Table 2.1. *Officer Candidate School (OCS) Beginning-of-Class (BOC) Sample Sizes and End-of-Class (EOC) Response Rates for the SelectOCS Phase 2 and 3 Samples*

Session	Phase 2			Phase 3		
	Class	BOC	EOC	Class	BOC	EOC
		<i>n</i>	<i>n</i> (%)		<i>n</i>	<i>n</i> (%)
1	2010-006	142	129 (90.8)	2011-010	107	78 (72.9)
2	2010-007	172	151 (87.8)	2011-011	121	84 (69.4)
3	2010-008	161	119 (73.9)	2011-012	103	42 (40.8)
4	2010-009	159	129 (81.1)	2011-013	60	48 (80.0)
5	2010-010	139	113 (81.3)	2011-014	68	52 (76.5)

Note. BOC = beginning of class; EOC = end of class. The response rate for Class 2011-012 was lower than that of other classes due to an administrative error that caused the EOC survey to be administered later than the target date.

After the graduation date, we requested information on candidate performance during OCS from administrative records. We also obtained candidate AFQT scores from two Army administrative databases. This process was repeated for five classes in 2010 for Phase 2 and another five classes in 2011 for Phase 3. In the future, Phase 3 participants will be asked to participate in further data collections at the end of their technical training, or Basic Officer Leader Course B (BOLC B), and just before the end of their ADSO (see Allen & Young, 2012 for a description of these data collections).

Database Development

Consistent with Phase 2, we scanned the OBEF and demographic data from the BOC data collections and merged them to create a master file. We then obtained administrative data from the Military Integrated Resource System (MIRS) and Total Army Personnel Database (TAPDB). We extracted individual ASVAB scores from these two databases along with a few additional demographic variables. Upon completion of the EOC data collections, these data were also processed. Finally, we obtained information on candidate performance from OCS. We then merged all of these data into the master database. Once we merged the data, the next step was to clean them.⁷ We flagged an individual's responses if more than 10% of the items had system missing values or if there was zero, or nearly zero variance, in their responses.

Participants

Consistent with SelectOCS Phase 1 and Phase 2, we conducted most of the analyses separately on the in-service and enlistment option samples. We classified candidates who indicated that they were "an enlisted Army Soldier" on the demographics form as in-service, and those who indicated that they were "a civilian with no prior military service" as enlistment option. We classified candidates who indicated one of the remaining three options as "hybrids."⁸ We created the "hybrid" group to limit the potentially confounding influence of these individuals on the two central groups of interest.

Key demographics for the SelectOCS Phase 2 and Phase 3 samples are contained in Table 2.2. The most striking difference between the two samples is the distribution of pre-service accession options. In Phase 2, more than a quarter of the sample was in-service, but in Phase 3, in-service constituted less than 10% of the sample. The percentage of enlistment option candidates increased accordingly from 56% in Phase 2 to 72% in Phase 3. However, the distribution of demographics across the Phase 2 and Phase 3 samples was relatively consistent, although with slightly more white, non-Hispanic males. This is consistent with the increase in representation of the enlistment option sample, which also trends toward white, non-Hispanic. The Phase 3 group also had higher incidence of advanced degrees than the Phase 2 sample.

The SelectOCS Phase 2 and 3 descriptive statistics for the continuously scaled demographics are contained in Table 2.3. For the enlistment option sample, the descriptive statistics were consistent for the two phases. The Phase 3 in-service sample had higher AFQT scores, was younger, and had less time in service and deployment time on average than did the Phase 2 in-service sample. However, the Phase 3 sample is small, so it is difficult to ascertain whether the mean differences were due to policy changes or if it was an artifact of the sample.

⁷ We constructed the data cleaning flags such that each individual measure in the OBEF received a "usability" flag, where each case was marked as "usable" or "unusable." When a case was marked as "unusable," we flagged and excluded it from subsequent analyses. We did not remove any cases entirely from the dataset; rather, we removed data only for the flagged measure(s) within a case. For example, if an individual had usable RBI data but unusable LKT data, then the RBI data were included in subsequent analyses, but the LKT data were not.

⁸ The remaining three options were "an enlisted Service member from another branch of the Armed Services," "a civilian with prior military service," and "a civilian who was previously enrolled in ROTC or at West Point."

Table 2.2. *Officer Candidate School (OCS) Demographics for SelectOCS Phase 2 and 3 Samples*

Demographic	In-Service (Phase 2 <i>n</i> = 216) (Phase 3 <i>n</i> = 36)		Enlistment Option (Phase 2 <i>n</i> = 434) (Phase 3 <i>n</i> = 331)		Total (Phase 2 <i>N</i> = 781) (Phase 3 <i>N</i> = 459)	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
	<u>Phase 2</u>					
Gender						
Male	166	76.9	360	82.9	628	80.4
Female	34	15.7	55	12.7	102	13.1
Ethnicity						
Hispanic	38	17.6	43	9.9	102	13.1
Non-Hispanic	156	72.2	368	84.8	615	78.7
Ethnicity						
White	122	56.5	337	77.6	539	69.0
American Indian/Alaskan	0	0.0	2	0.5	2	0.3
Asian	10	4.6	22	5.1	42	5.4
Black or African American	61	28.2	27	6.2	102	13.1
Pacific Islander	1	0.5	1	0.2	2	0.3
Multiple	9	4.2	32	7.4	50	6.4
Education						
Some college	53	24.5	0	0.0	57	7.3
Bachelor's degree	114	52.8	355	82.8	553	71.3
Some graduate school	30	13.9	22	5.1	68	8.8
Master's degree	18	8.3	41	9.6	72	9.3
Doctorate or equivalent	1	0.5	9	2.1	14	1.8
Gender						
Male	25	69.4	277	83.7	382	83.2
Female	7	19.4	40	12.1	54	11.8
Ethnicity						
Hispanic	4	11.1	21	6.3	37	8.1
Non-Hispanic	29	80.6	287	86.7	390	85.0
Race						
White	22	61.1	260	78.5	350	76.3
American Indian/Alaskan	2	5.6	0	0.0	5	1.1
Asian	1	2.8	19	5.7	24	5.2
Black or African American	9	25.0	25	7.6	42	9.2
Pacific Islander	0	0.0	1	0.3	1	0.2
Multiple	0	0.0	19	5.7	25	5.4
Education						
Some college	3	8.3	0	0.00	3	0.7
Bachelor's degree	26	72.2	257	77.6	357	77.8
Some graduate school	2	5.6	25	7.6	34	7.4
Master's degree	5	13.9	43	13.0	57	12.4
Doctorate or equivalent	0	0.0	6	1.8	8	1.7

Note. In-Service: Candidates who were Army Soldiers prior to OCS; Enlistment Option: Candidates who were civilians with no military service prior to OCS. Hybrids and candidates who did not answer the pre-service question were not included as separate groups in this table. “Hybrids” consist of candidates who came through the enlistment option, but had some prior military exposure (service from another military branch, prior military service, or experience from West Point or ROTC). We included Hybrid data, as well as data from those who did not answer the pre-service question, in the “Total” column. Percentages do not add up to 100% within each demographic variable due to missing data.

Table 2.3. *Officer Candidate School (OCS) Continuous Demographics for SelectOCS Phase 2 and 3 Samples*

	In-Service			Enlistment Option			Total		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
					<u>Phase 2</u>				
AFQT	188	71.69	18.80	435	88.44	9.02	688	83.14	14.67
Age	209	31.57	3.89	422	26.81	3.69	745	28.78	4.55
Time in Service (months)	214	86.94	48.11	436	N/A	N/A	769	34.31	49.61
Time Deployed (months)	214	16.96	30.17	434	N/A	N/A	765	5.74	18.20
					<u>Phase 3</u>				
AFQT	36	75.72	17.51	326	88.79	8.39	452	86.70	10.76
Age	34	29.00	3.26	320	25.88	3.40	440	26.60	3.72
Time in Service (months)	36	73.08	50.03	326	N/A	N/A	454	16.13	32.94
Time Deployed (months)	36	13.36	15.68	325	N/A	N/A	452	2.28	7.14

Note. In-Service = Candidates who were Army Soldiers prior to OCS; Enlistment Option = Candidates who were civilians with no military service prior to OCS. N/A = not applicable. Hybrids and candidates who did not answer the pre-service question were not included as separate groups in this table. “Hybrids” consist of candidates who came through the enlistment option, but had some prior military exposure (service from another military branch, prior military service, or experience from West Point or ROTC). We included Hybrid data, as well as data from those who did not answer the pre-service question, in the “Total” column.

These results continue a number of trends found when comparing the Phase 1 and Phase 2 samples (Russell et al., 2011; Russell & Tremble, 2011). First, the results demonstrate that the OCS mission is decreasing. At the time of this writing, combat operations had recently ended for Operation Iraqi Freedom and were drawing down for Operation Enduring Freedom. Consequently, the Army is also reducing the size of its force and no longer needs to commission as many officers as in 2008 when this research began. Second, the results demonstrate a continued trend favoring enlistment option candidates over in-service candidates for OCS. This may be due in part to new selection requirements set in October of 2010 that (a) require all candidates to have a 4-year degree from an accredited college prior to entering OCS and (b) suspend waivers for time in service, age, medical, and moral reasons (MilPer Message 10-164; see also Tice, 2010). Third, the range of AFQT scores in both samples continues to become more restricted compared to population estimates (Bureau of Labor Statistics, 2005); suggesting predictive validity analyses that include raw AFQT scores will underestimate its ability to predict key outcomes. Overall, these results suggest that the OCS selection and accessioning process is changing.

Predictor Instruments

The OBEF and AFQT served as the primary instruments for assessing the predictor space. The SelectOCS Phase 3 OBEF was very similar to the SelectOCS Phase 2 OBEF (Russell et al., 2011). In Phase 2, the OBEF comprised five instruments: (a) RBI, (b) TAPAS, (c) Work Values (ranked and scaled), (d) LKT, and (e) Army Identification. Certain attitudinal scales (e.g., Affective Commitment, Career Intentions) described in the criterion section were also administered with the OBEF. The Phase 3 OBEF comprised the same five instruments; however, we added five scales to the RBI and dropped one scale from the Work Values. We describe these changes in more detail in the appropriate sections below.

Armed Forces Qualification Test (AFQT)

Candidates' AFQT scores were drawn from administrative records. For each candidate in our sample, we extracted ASVAB scores from MIRS, which contains information about Army applicants, and the TAP-DB, which provides information on Soldiers from the time they enlist until they separate. Because we want to generalize these results to an applicant population, the MIRS ASVAB scores took precedence over the TAP-DB scores in determining each candidate's AFQT score. In instances where a MIRS record could not be identified for a candidate in our sample (e.g., some in-service candidates applied to the Army many years ago, so their records were difficult to obtain), we substituted the TAP-DB scores.

Rational Biodata Inventory (RBI)

The RBI is a non-cognitive self-report instrument measuring aspects of individual personality and motivational attributes. The items ask respondents to think of their previous experiences in a particular area relevant to the latent construct of interest (e.g., Stress Tolerance) and respond on a 1 (low) to 5 (high) scale. The points on the scale change depending on the nature of the question. For example, some items ask how often the respondent had a particular experience, in which case the scale ranged from 1 (never) to 5 (very often), whereas other items

ask respondents to indicate the extent to which they agree with the statement on a scale that ranged from 1 (strongly disagree) to 5 (strongly agree).

We chose scales measuring latent constructs thought to relate to officer performance and retention for inclusion in the Phase 2 RBI (see Table 2.4). In Phase 3, we added five scales (33 items) to the RBI administered in Phase 2. Their inclusion was based on previous ARI research, including enlisted Soldier selection research (Knapp & Heffner, 2010; Narcissism), officer job analysis findings (Paullin et al., 2011; Emotional Stability), officer predictor and criterion development work (Paullin et al., 2012; Goal Expectations), and the SelectOCS Phase 1 experimental scales (Russell & Tremble, 2011; Tolerance for Ambiguity and Social Acuity).⁹

Similar to SelectOCS Phase 2, most of the RBI scales in Phase 3 exhibited adequate reliability; with coefficient alpha reliability estimates ranging from .54 to .78 with a mean of .69 (see Table B.1, Appendix B). The scales with coefficients below .60 were Achievement ($\alpha = .54$) and Hostility to Authority ($\alpha = .58$). Due to heterogeneity in the item content, we expected the biodata scales to have lower overall internal consistency reliability than traditional self-report measures (Kilcullen, White, Mumford, & Mack, 1995). Additionally, the means and standard deviations had acceptable ranges for prediction, although the means for some scales (Achievement, Generalized Self-Efficacy, Leader Self-Efficacy, and Affective Commitment) were above 4.0 on a 5-point scale. Because the variance on these scales is restricted, they might be less predictive of key outcomes than are other RBI scales.

Tailored Adaptive Personality Assessment System (TAPAS)

The Tailored Adaptive Personality Assessment System (TAPAS; Drasgow, Stark, & Chernyshenko, 2006; Stark, Drasgow, & Chernyshenko, 2008) is an item response theory-based computerized adaptive personality test designed to measure up to 22 lower-order facets of the five-factor model of personality (Costa & McCrae, 1992). The version administered in this effort was the same non-adaptive, paper-based, 12-dimension measure with 96 paired-response items used in SelectOCS Phase 2 (see Table 2.5). TAPAS uses multidimensional pairwise preference (MDPP) personality items scored using item response theory (IRT) methods. The MDPP format is designed to be more faking-resistant by pairing items that have similar levels of social desirability. This makes it more difficult for respondents to “fake good” because both options in the pairing have approximately equal positive valence (for complete scoring information, see Drasgow et al., 2006).

⁹ The RBI also includes a Response Distortion scale designed to catch individuals who are responding to the RBI in a socially desirable manner. In SelectOCS Phase 1, the Response Distortion scale was found to be modestly correlated with other RBI scales, ranging from a low of .01 to a high of .33 (absolute values; Russell & Tremble, 2011). Adjusting scores for response distortion was out of scope for the present analysis.

Table 2.4. *RBI Scales Administered in Phase 3*

Scale	Items	Definition
<i>From Phase 1:</i>		
Achievement Orientation	6	The willingness to give one's best effort and to work hard towards achieving difficult objectives.
Affective Commitment	9	The extent to which a candidate feels emotionally attached to the Army.
Fitness Motivation	8	Degree of enjoyment from participating in physical exercise. Willingness to put in the time and effort to maintain good physical conditioning.
Hostility to Authority	6	Being suspicious of the motives and actions of legitimate authority figures. Viewing rules, regulations, and directives from higher authority as punitive and illegitimate.
Learning Self-Efficacy	6	Belief in one's ability to discern what might be on tests, examinations, or other demonstrations of ability.
Peer Leadership	6	Seeks positions of authority and influence. Comfortable with being in charge of a group. Willing to make tough decisions and accept responsibility for the group's performance.
Generalized Self-Efficacy	6	Feeling that one has successfully overcome work obstacles in the past and that one will continue to do so in the future.
Stress Tolerance	9	Ability to maintain one's composure under pressure. Remaining calm and in control of one's emotions instead of feeling anxious and worried.
<i>Added in Phase 2:</i>		
Interest in Leadership	4	Preference for serving as a leader, being in a position of influence on project teams in which one serves.
Leader Self-Efficacy	8	Belief in one's ability to successful lead groups.
Tolerance for Injury	6	Tolerance for situations where risk is possible. Attraction to activities involving risk.
Equity Sensitivity	6	Degree to which one is sensitive to contributing his/her efforts in comparison to the efforts contributed by others.
ARC Hostility to Authority	5	Belief that superiors abuse their power and take advantage of their employees. This is a subdimension of the higher-order scale Assessment of Right Conduct (ARC).
<i>Added in Phase 3:</i>		
Emotional Stability	5	The extent to which a candidate is resilient to the stresses of everyday life.
Goal Expectations	6	Has a high expectation of future achievements relative to other officers.
Narcissism	6	Being excessively preoccupied with satisfying one's own needs and desires.
Tolerance for Ambiguity	8	Ability to tolerate work situations where the right goal or the correct path to the goal is vague and ill-defined.
Social Acuity	8	The ability to understand the feelings/motives of others and the ability to take this information into account and respond appropriately in interactions.

Note. Items = final number of items scored for each scale. The RBI also has a "Response Distortion" scale designed to detect socially desirable responding. The RBI scale scores were not adjusted using this scale, and this scale was not included in any subsequent analyses.

Table 2.5. *TAPAS Scales by Big Five Trait*

Scale	Description	Big Five Trait
Dominance/Leadership	High-scoring individuals are domineering, take charge, and are often called “natural leaders” by their peers.	Extraversion
Trust/Cooperation	Individuals scoring high on this facet are trusting, cordial, cooperative, and easy to live with.	Agreeableness
Optimism	Individuals with high scores on this factor are described as happy and able to maintain a positive outlook.	Emotional Stability
Achievement	Individuals with high scores on this factor are described as hard working, ambitious, confident, and resourceful.	Conscientiousness
Non-Delinquency	People with high scores on this facet tend to comply with current rules, customs, norms, and expectations. They dislike changes, and do not challenge authority.	Conscientiousness
Responsibility	Those scoring high on this facet express willingness to demonstrate personal responsibility and dedication to duty.	Conscientiousness
Even Temper	Those scoring low on this facet have a tendency to experience a range of negative emotions including irritability, anger, hostility, or aggression; those scoring high tend to be calm and stable.	Emotional Stability
Adjustment	Those scoring high on this facet demonstrate flexibility in behavior and ability to overcome setbacks quickly.	Emotional Stability
Intellectual Efficiency	Individuals with high scores on this factor are able to process information quickly and would be described by others as knowledgeable, astute, and intellectual.	Openness to Experience
Curiosity/Continuous Learning	Individuals with high scores on this facet are characterized as inquisitive and perceptive; they are interested in experimenting with objects and substances.	Openness to Experience
Tolerance	Individuals scoring high on this facet like to attend cultural events or meet and befriend people with different views; they adapt better to novel situations.	Openness to Experience
Physical Conditioning	High scoring individuals routinely participate in vigorous sports or exercise and enjoy hard physical work.	N/A

Note. Adapted from Stark et al., 2008.

The means and standard deviations across the scales in the present effort (Table B.1, Appendix B) are similar to those found in Phase 2 and other ARI-sponsored projects, where it has shown promising predictive efficacy (Knapp & Heffner, 2010; Russell et al., 2011). Because of the scoring algorithm and response format, we could not compute traditional alpha reliabilities for the TAPAS.

Work Values

As with SelectOCS Phase 2, we used two methods to assess each candidate’s work values. The first was the rank-ordered scales originally developed as part of ROTC Cadet Background Experiences Form (CBEF; Putka, 2009). In this method, we asked candidates to rank-order a set of values statements and then indicate which of the ranked statements would

need to be present for the respondent to consider the job ideal. The algorithm used to score these instruments has been used to score other ARI-sponsored work values instruments (e.g., Knapp, Sager, & Tremble, 2005), and is described in more detail elsewhere (e.g., see Appendix C in McCloy et al., 1999). We derived scores for 11 work values using this method (see Table 2.6).

The second method used to assess a candidate's work values was a traditional single-item with 5-point Likert-scale response format. We assessed two important latent factors not tapped in the rank-ordered set using this method: (a) Benevolence and (b) Social Work Environment (Table 2.6). We dropped a third dimension administered in the SelectOCS Phase 2 OBEF (i.e., self-development) due to poor psychometrics and to free up administration time for the new RBI scales. We administered these scales with the RBI items in the SelectOCS Phase 3 OBEF.

Table 2.6. *Phase 3 Work Values Content*

Scale/Item	Definition	
<i>Work Values (Rank-Order)</i>		
Challenge		Doing work that is challenging.
Comfort		Working in a comfortable, relaxed environment.
Home		Doing work that keeps one close to home.
Leadership		Providing guidance and direction to others.
Pay		Receiving a good salary and benefits.
Recognition		The desire to receive recognition or praise for what one does.
Self-Direction		The ability to determine one’s own way to do tasks.
Selfless Service		The willingness to contribute to society and the well-being of others.
Structure		The desire for having well-defined rules for accomplishing tasks.
Teamwork		The desire to work as part of a team.
Variety		The desire to work on a variety of types of problems.
<i>Work Values (Likert Scale)</i>	<i># of Items</i>	<i>Definition</i>
Benevolence	5	Preference for work that helps others and makes the world a better place.
Social Work Environment	5	Preference for work that involves close relationships with others.

The Phase 3 coefficient alpha reliability estimates for the two Likert-scaled work values were acceptable at around .70 (see Appendix B, Table B.1). The remaining descriptive statistics were similar to those found in SelectOCS Phase 2 (Russell et al., 2011).

Leader Knowledge Test (LKT)

The LKT presented respondents with a list of 30 traits and 30 skills (derived from leadership and personality literature) and instructed them to rate the importance of each trait or skill to performing successfully as a company grade leader. They were asked to respond on a 1 (Not at all important) to 10 (Extremely important) scale. In SelectOCS Phase 3, we scored the LKT in the same manner as in Phase 2—by comparing the pattern of importance ratings provided by the candidate to a “key” of ratings provided by captains in the Captain's Career Course (CCC) (Allen et al., 2011). In other words, we treated the ratings of importance provided

by the captains as the “correct” answers for the purposes of this analysis. We then constructed scores for each candidate by (a) standardizing the item-level data within participant (for both captains and the Phase 3 officer candidates), (b) computing a response “key” using the captains’ means, (c) subtracting each SelectOCS candidate’s response from the keyed response, and (d) subtracting that difference from the integer 10. We computed separate scores for the trait and skill lists. More details about this scoring procedure can be found in Pearlman, Allen, Putka, Hooper, and Waters (2009) and McDaniel, Psotka, and Legree (2009). Estimates of internal consistency reported in Table B.1 of Appendix B were acceptable for the traits scale ($\alpha = .78$) but low for the skills scale ($\alpha = .56$), although low coefficient alpha values are not unexpected when using the standardized scoring approach (Pearlman et al., 2009).

Army Identity Structure

Three dimensions of Army Identity Structure (Overlap, Concept, and Conflict) were measured using three single-item measures. The single-item measures were graphical representations of the constructs of interest, an approach that was adapted from Shamir and Kark (2004) in SelectOCS Phase 1 (Robbins et al., 2011). Each item was scaled on a 1 (low) to 7 (high) scale. Because these were single-item measures, we could not assess internal consistency. However, as shown in Table B.1 in Appendix B, there was enough variance in the scores to make these structure items useful for prediction.

Criterion Measures

As described in Chapter 1, we are interested in the extent to which the OBEF predicts performance and continuance criteria. We defined the criterion space using participating candidates’ responses on the EOC measure and their performance during training, derived from administrative records at OCS.

End-of-Class (EOC) Measure

One criterion of particular interest to the Army is whether an officer stays beyond his or her initial ADSO. However, to assess this directly, we would need to wait more than three years, when every officer candidate in our sample will reach the end of their ADSO (see Strickland, 2005, for such a research project of enlisted Army personnel). Another criterion of particular interest is candidates’ performance as a leader when in their unit of assignment. As with separations, this would mean a delay in the collection of performance information in addition to logistical difficulties (e.g., lack of administrative records) in collecting such data. Although collecting in-unit performance and long-term continuance are planned in future data collections, we can learn much about individuals’ separation intentions and performance early in their term of service by using proximal antecedents of these criteria.

Accordingly, the web-based EOC survey included two types of items: (a) attitudinal items, and (b) peer evaluations. The attitudinal items are the same as those administered in SelectOCS Phases 1 and 2 and the peer ratings are the same as those administered in Phase 2. The attitudinal dimensions assessed in the EOC measure are described in Table 2.7. We also asked candidates to report the branch to which they were assigned and if they were satisfied with this assignment. As with the OBEF measures, data from a candidate were flagged as unusable if

more than 10% of responses to the attitudinal items were missing, or if he/she completed the entire EOC instrument in less than 2 minutes (as calculated by the surveying software used to collect the EOC data). As in Phase 2, the coefficient alpha reliability estimates for items with more than one component item were high ($\alpha = .82$ and $.87$ for Continuance and Affective Commitment, respectively). In the present analysis, we believe the Affective Commitment and Career Intentions scales will serve as particularly good criteria for assessing officer continuance, because previous research has found both to be indicators of separation in previous analyses (e.g., Hom, 2011; Strickland, 2005).

Table 2.7. *End-of-Class Criterion Variables*

Scale	Description
Continuance Commitment	Four-item scale measuring the extent to which external factors affect candidates' commitment to complete their current terms of service. An example item is "It would be too costly for me to leave the Army in the near future." Items were scored on a 1 to 5 scale ranging from "Strongly Disagree" to "Strongly Agree."
Affective Commitment	Four-item scale measuring the extent to which a candidate felt emotionally attached to the Army. An example item is "I feel like 'part of the family' in the Army." Items were scored on a 1 to 5 scale ranging from "Strongly Disagree" to "Strongly Agree."
Morale	A single-item measure of a candidate's current level of morale (i.e., "What is your current level of morale?"). The item was scored on a 1 to 5 scale ranging from "Very Low" to "Very High."
Branch Satisfaction	A single-item measure of a candidate's satisfaction with branch assignment. The item was scored on a 1 to 5 scale ranging from "I am very dissatisfied with my branch assignment" to "I am very satisfied with my branch assignment."
Career Intentions	A single-item measure of a candidate's active duty career intentions (i.e., "What are your current active duty career intentions?"). The item was scored on a 1 to 5 scale ranging from "I will definitely quit the Army upon completion of my obligation" to "I plan to stay in the Army beyond 20 years or until retirement."
Leadership Potential	A two-item measure of a candidate's potential for company-grade and field-grade leadership. Items were scored on a 1 to 5 scale ranging from "Likely to be a poor or marginal performer" to 5 "Likely to be a truly exceptional performer." Ratings were provided by the target candidate's peers.

Candidates were also asked to evaluate their own potential and that of their squad members (i.e., those with whom they train most closely during OCS) on their likely effectiveness in a company grade leadership position (i.e., as a platoon leader or company commander directing relatively small units to achieve clear, immediate, and well-defined goals) and a field grade leadership position (i.e., as a battalion commander or brigade commander directing larger organizations and broader systems to achieve more complex long-term goals). For both scales, respondents were asked to provide their ratings on a 1 ("Likely to be a poor to marginal performer") to 8 ("Likely to be a truly exceptional performer") scale. Respondents' ratings data were marked as unusable if (a) it took them less than 2 minutes to complete the entire EOC survey or (b) if they rated more than five people and assigned the same rating on both scales to all of their ratees. For each candidate, we computed a mean company-grade leader score (across peers) and a mean field-grade leader score (across peers). As in Phase 2, we found that the two scores were highly correlated ($r = .95$ in the Phase 3 sample), and therefore computed a peer

leadership composite score by averaging the two scores. Self-ratings were not included in the peer leadership scores.¹⁰

As with the attitudinal criteria, we believe that these Peer Ratings of Leadership potential are good proxies of future leadership performance. Previous work by Kraut (1975) found peer ratings obtained in training to be good predictors of manager performance, as assessed by promotions and performance appraisals. Correlations were mainly in the .30s and .40s. Furthermore, he found that the predictive efficacy of these peer ratings held up more than 20 years later (Kraut, 2005).¹¹

The descriptive statistics and intercorrelations for these measures in the Phase 3 sample can be found in Appendix B, Table B.2. As expected, they exhibited acceptable psychometric properties for further analysis, with coefficient alpha reliability estimates ranging from .82 to .87 for the attitudinal scales. The interrater reliability estimates were also acceptable, with k -rater reliabilities of .76 and .74 for the company-grade and field-grade ratings, respectively.¹² The patterns of intercorrelations were also consistent with theoretical expectations.

OCS Performance Data

At the end of OCS, officer candidates are rank-ordered according to their performance during the 12-week course. This ranking, called the Order of Merit List (OML), is a combination of scores in three performance areas: (a) academic performance, (b) leadership, and (c) physical fitness. Candidates receive scores in these three areas through written and physical tests, performance during field training exercises, and the collection of cadre and peer ratings. The following summarizes the most pertinent metrics gathered during OCS.

1. Academic Examinations — Candidates complete eight academic tests. The passing requirement is 70% on each. If they fail, candidates can retest once; only three total retests (on the eight tests) are allowed. Candidates are tested on the following areas: (a) tactics and operations, (b) call for fire, (c) history (two tests), (d) supply, (e) training management, (f) military intelligence, and (g) leadership, justice, and ethics.
2. Leadership — Candidates serve in a number of leadership roles (from Team Leader to Company Commander) both in garrison and during field leadership exercises (FLX). Instructors and cadre conduct evaluations for leadership occurrences. Candidates must receive a minimal “Satisfactory” rating on 50% of their leadership evaluations, called the Leader Performance Evaluation Report (LEPR). Candidates receive points for each “Satisfactory” and “Excellent” rating on the LEPR, but no points for a “Needs Improvement” rating. Candidates are evaluated on their demonstrations of (a) Army Values, (b) leadership attributes (emotional, mental,

¹⁰ The correlation between the peer and self-ratings ranged from .22 to .23 in the combined Phase 2 and Phase 3 sample.

¹¹ The authors would like to thank Robert Kaiser for providing these references.

¹² Interrater reliability estimates were computed using $G(q,k)$ (Putka, Le, McCloy, & Diaz, 2008), where k is equal to the harmonic mean of the number of raters per ratee ($k = 4.92$). The corresponding single-rater coefficients (i.e., $G[q,1]$) were .45 and .42 for the company-grade and field-grade ratings, respectively.

physical), (c) leadership skills (interpersonal, conceptual, technical, tactical), and (d) leadership actions (influencing, improving, operating).

3. **Physical Fitness** — Candidates must receive a passing score on a variety of physical fitness evaluations, including two scores of at least 60 on each portion of the APFT, runs of multiple distances, foot marches with full gear of multiple distances, and an obstacle course.

In previous SelectOCS phases, there was some class-specific variation in how the OML was computed. Procedures implemented at OCS between Phase 2 and Phase 3 ensures that the OML is computed the same way for each class. However, the scores for particular activities (e.g., LEPR ratings) that feed into the OML are computed differently depending on the class. On the other hand, we can be reasonably confident, given findings in previous research (Oliver et al., 2011), that certain aspects of the OML are scored consistently across classes. For example, the APFT is scored the same way based on standards set by Army Field Manual 21-20. Another example is the OCS History courses, which OCS has scored similarly for years. We could also potentially use these consistently scored events as criteria for further analysis. We describe the variables considered for further analysis in more detail in Table 2.8. We were unable to obtain OML data for one of the Phase 3 classes in time for data analysis.

Table 2.8. *OCS Administrative Criterion Variables*

Scale	Description
Academic Performance	Archival score maintained by OCS that was a unit-weighted sum of the candidates' academic course scores. The raw scores were standardized within class to a mean of 0 and a standard deviation of 1.
History Performance	Candidates completed two history courses during OCS. The combined raw scores for these two courses feed into candidates' overall OML. Each course is worth 100 points, and a candidate must receive a minimum of 70 to graduate. The score used for our analysis was an average of the scores from the two courses, which ranged from a low of 39 to a high of 100 in the present sample.
Physical Performance	Archival score maintained by OCS that was a unit-weighted sum of the candidates' last Army Physical Fitness Test (APFT) and their scores on fitness exercises. The raw scores were standardized within class to a mean of 0 and a standard deviation of 1.
Army Physical Fitness Test (APFT) Scores	Candidates completed the APFT three times during OCS. An arithmetic average of candidates' first and third APFT scores (the two APFTs for record during OCS that contribute to the final OML) was used as an additional criterion in the present effort. Total APFT scores range from 0 to 300, although scores in our sample ranged from 204 to 300.
Leadership Performance	Archival score maintained by OCS that was a unit-weighted sum of the candidates' leadership exercise scores and peer/trainer ratings of leadership. The raw scores were standardized within class to a mean of 0 and a standard deviation of 1.
Total OCS Performance Score	The raw Total OCS Performance Scores were standardized within class to a mean of 0 and a standard deviation of 1.

Subgroup Differences on Predictors

Large subgroup differences on the OBEF predictor instruments can lead to adverse consequences for the Army generally and OCS specifically. Predictors with large subgroup differences have higher potential for selecting out protected minority groups (i.e., women, Blacks, and Hispanics in this analysis), which can lead to their systematic underrepresentation. A scenario where the OBEF reduces (or at least does not increase) adverse impact beyond the Army's current selection instrument – the ASVAB, as determined by AFQT scores in this analysis – provides additional justification beyond predictive efficacy for its operational use. To examine subgroup differences, we first computed the means and standard deviations for each OBEF scale on the subgroups of interest (i.e., gender, race, and ethnicity). We also conducted this analysis by pre-service accession option (in-service versus enlistment option). We then determined the extent to which the means differ by using an independent samples *t*-test and a Cohen's *d* (Cohen, 1988). For interpreting Cohen's *d*, we followed the rule of thumb stating that a value of less than 0.30 is considered "small," 0.50 "medium," and 0.80 "large." The results of this analysis are summarized below (complete results can be found in Table B.3 in Appendix B). In our analysis, we calculated all *d* values using a "minority group – majority group" approach, such that a positive *d* indicates the mean was higher for the minority group (female, Black, Hispanic, enlistment option) than for the majority group (male, White, non-Hispanic, in-service option).

In terms of pre-service option, candidates who entered OCS through the enlistment option had significantly higher AFQT scores than candidates who entered through the in-service option ($d = 0.75$), despite the requirement that all applicants have a 4-year degree. While the sample size of the in-service group was relatively small, this finding was also reported in Phases 1 and 2, where the sample sizes were much larger. The differences between the two groups on the OBEF scales were generally small and not statistically significant. The exception to this was the RBI Tolerance for Injury scale, where candidates entering OCS through the enlistment option ($M = 3.82$; $SD = 0.61$) scored significantly higher than candidates entering OCS through the in-service option ($M = 3.54$; $SD = 0.65$; $d = 0.43$). However, previous SelectOCS projects have shown that while the means and standard deviations for the two service options are similar, the scales that predict key outcomes of interest are different between the two groups (Russell & Tremble, 2011; Russell et al., 2011). Therefore, subsequent analyses will continue to separate those two groups.

Unlike the pre-service option results, there were a number of large and statistically significant differences between males and females on the OBEF scales. However, many of these differences favored the female candidates. Specifically, the RBI Peer Leadership ($d = 0.47$), RBI Achievement ($d = 0.44$), RBI Generalized Self-Efficacy ($d = 0.29$), RBI Hostility to Authority ($d = -0.35$), LKT ($d = 0.24$), Leadership Work Value ($d = 0.32$), TAPAS Tolerance ($d = 0.58$), RBI Social Acuity ($d = 0.32$), and TAPAS Leadership ($d = 0.28$) scales all had significantly higher (or lower in the case of the negatively valenced Hostility to Authority scale) means for females than males. This suggests that if these scales are predictive of key outcomes and were included in a prediction composite, they could counterbalance the negative effect of other scales included in the composite that favor males. The OBEF scales that favor males include the RBI Fitness Motivation ($d = -0.65$), RBI Tolerance for Injury ($d = -0.46$), and TAPAS Adjustment ($d = -0.55$).

scales. Regardless of the direction, the measures with the largest gender differences were the RBI (Average $|d| = 0.25$) and the TAPAS (Average $|d| = 0.21$).

For race (Black versus White), there were large subgroup differences on the AFQT favoring the majority group (Race $d = -0.85$). Comparatively, there were few significant differences for the OBEF scales, and those observed tended to favor the minority group. Specifically, Black candidates scored lower than White candidates did on the RBI ARC Hostility to Authority scale ($d = -0.43$) and higher than White candidates on the TAPAS Tolerance ($d = 0.33$), TAPAS Optimism ($d = 0.44$), TAPAS Adjustment ($d = 0.49$), and Army Identity Structure Overlap ($d = 0.36$) scales. Sample sizes of ethnic groups were too small to permit subgroup comparisons ($n = 12-14$ with data on the OBEF scales).

In summary, these results suggest that the OBEF in general has much smaller pre-service option, and race subgroup differences than the AFQT. The differences that do exist tend to be small or to favor the minority subgroup. By contrast, the OBEF subgroup differences for gender tended to be much larger than the difference found for the AFQT. However, some of these differences favored males and others favored females. This suggests that, in a selection context, the component scales could potentially be balanced so that there is no adverse impact to women, depending on the predictive efficacy of the component scales.

Summary

The purpose of this chapter was to describe the SelectOCS Phase 3 research method. We began by describing the research design and the procedures used to collect the data. We then described the steps taken to clean the data and construct the analysis database, followed by some initial results, including the characteristics of the sample, the psychometric characteristics of the predictors and criteria, and the subgroup differences on the predictor measures. All OBEF components (RBI, TAPAS, LKT, Army Identity Structure, and Work Values) have acceptable psychometric characteristics to warrant further analysis.

Chapter 3: Predicting Candidate Performance with the OBEF

This chapter describes the analyses conducted to validate the OBEF in an independent sample of officer candidates. The first objective of these analyses was to examine the differences in incremental validity between the SelectOCS Phase 2 and Phase 3 samples of candidates. In doing so, we determined whether the OBEF was functioning similarly in an independent sample of candidates, with the goal of cross-validating the Phase 2 findings. If the Phase 2 and Phase 3 results are similar, we can then combine the two samples for subsequent analyses. Given the Phase 3 results were limited by small sample sizes, combining the two samples provides greater power and confidence in the results of the analyses. The second objective was to select the most promising individual measure within the OBEF for predicting officer performance and continuance. The chosen measure was the base for the composite formation analyses described in Chapter 4.

Analytic Approach

In the SelectOCS Phase 1 and Phase 2 projects, the potential for each OBEF measure to contribute to the Army's current selection procedures was determined by examining the incremental validity of each measure beyond a baseline measure, in this case, AFQT (Russell & Tremble, 2011). We used the same basic procedures in the current effort, with a few of modifications. Most notably, we used missing data analyses to account for missing data in the criteria. This enabled us to retain data we would have otherwise lost had we used the same analytic procedures used in Phases 1 and 2. We used the following procedures to examine the predictive efficacy of the experimental measures:

Differential Prediction. The purpose of the analysis was to cross-validate the Phase 2 criterion-related validity analyses with the Phase 3 data. We carried out this analysis with the following steps:

1. Examine subgroup differences between the two samples, using the same procedures as the subgroup difference analyses described in Chapter 2.
2. Conduct multiple group measurement equivalence analyses to compare the incremental validity of the two samples.
3. Determine the equivalence of the incremental validity estimates for the two samples.
4. Compare the bivariate correlations.

Incremental validity. The purpose of the incremental validity analysis was to examine the criterion-related validity of each OBEF instrument to select a "best bet" measure for composite formation. We carried out these analyses in the following steps:

1. Compute the incremental validity using Ordinary Least Squares (OLS) regression with Full Information Maximum Likelihood (FIML) missing data analyses.
2. Determine the statistical significance of the incremental validity of the OBEF measures beyond AFQT.
3. Estimate a variance/covariance matrix using FIML missing data analyses.
4. Correct the variance/covariance matrix for direct range restriction on AFQT.

5. Compute incremental validity estimates using the corrected correlation matrix.
6. Correct the incremental validity estimates for shrinkage.

We describe these steps in more detail below.

As described in Chapter 2, one limitation of the current data is there were very few candidates in our sample who entered OCS through the in-service option. Consequently, we did not conduct the differential prediction analyses for the in-service sample and were unable to cross-validate the in-service Phase 2 validity results using the Phase 3 sample. However, we did combine the Phase 2 and Phase 3 in-service samples and examine the incremental validity of the predictor measures.

Missing Data Estimation

Studies have shown that missing data can lead to low power and downwardly biased estimates of model parameters (Enders & Bandalos, 2001; Muthén, Kaplan, & Hollis, 1987; Roth, 1994; Schafer & Graham, 2002). As is often the case when data are collected at two time points, we were unable to capture criterion data for all of the individuals who took the OBEF survey. Of the 459 candidates who took the OBEF survey, 304 (66.2%) responded to the EOC survey. In addition, we were able to collect OML data on only 315 (68.6%) of the candidates. Application of traditional methods to deal with missing data (e.g., listwise deletion) would result in small samples, potentially biased estimates, and limitations on the interpretability of the results.

To deal with the missing data, we opted to conduct regression analyses using full information maximum likelihood (FIML) missing data estimation. Studies have shown that FIML estimation produces parameters that are less biased than those produced by multiple imputation or that result from listwise, or pairwise deletion (Enders, 2001). FIML uses all of the available data to estimate the likelihood value of the parameter estimates for each individual. Specifically, when there are missing observations, matrices are produced that incorporate both observed and missing parameter information. The information associated with the complete portion of the vector is used to estimate the likelihood value of the variables with missing data (Enders, 2006). Note that unlike other missing data estimation methods, data are *not* imputed into the database. FIML accounts for missing data by including additional information to estimate the parameters in the analysis.

We conducted missing data analyses using the FIML feature in MPLUS (Muthén & Muthén, 2004). The final Phase 3 analysis sample included 33 in-service officer candidates and 319 enlistment option officer candidates. The Phase 2 analyses sample included 208 in-service officer candidates and 429 enlistment option officer candidates. We applied missing data analyses to both the differential prediction analyses and the incremental validity analyses.

Differential Prediction

We conducted a number of analyses to compare the Phase 2 and Phase 3 results. First, we examined the mean differences in the predictor and criterion scores between the Phase 2 and

Phase 3 cohorts. We computed Cohen's d (Cohen, 1988), an effect size expressed in standard deviation units, to estimate cohort differences on the predictors.

Next, we conducted multiple-group measurement equivalence analyses (or simply "multi-group analysis") to compare the incremental validity of the two samples. Multi-group analysis is used to examine multiple populations in a single data set and can be used to assess the measurement equivalence of the groups' parameter estimates. When conducting multi-group analyses within the same overall model, separate covariance matrices and model parameters are estimated for each group. Using this method, we can test a model where the parameter estimates of the different groups can be constrained and compared to a model where the parameter estimates are free to vary across groups.

For the current analysis, an FIML regression model was specified for each group, and regression coefficients were estimated simultaneously. We examined different models for each predictor/criterion relation. AFQT was entered in the model first, followed by the predictor set. Two models were estimated for each relation. First, the regression parameters were freely estimated; that is, different estimates were produced for each group. Then a second model was estimated where the regression parameters were constrained to be equal across groups. Overall model fit indices for the model with freely estimated parameters were compared to the fit indices of the model with constrained parameters using a chi-square difference test. If the models are significantly different, where the model with the unconstrained estimate fits significantly better, then we can conclude there are differences between the two groups' parameter estimates. If there is no difference in model fit between the unconstrained and the constrained models, the more parsimonious model is considered to be better fitting. In the present analyses, this would indicate that the Phase 2 and Phase 3 parameter estimates could be considered to be equal, thus cross-validating the Phase 2 results.

Finally, we examined the bivariate correlations between the predictor scales and the criteria for the Phase 2 and Phase 3 samples to determine if there were discrepancies in the relations and to identify specific scales that differed across the two samples.

Incremental Validity

To assess the criterion-related validity of the OBEF, we conducted OLS regression analyses to examine the incremental validity of each predictor set (i.e., RBI, TAPAS, Work Values, Army Identification, and LKT) over AFQT. Incremental validity is an estimate of the change in the multiple correlation (ΔR) when a new predictor or set of predictors is added to a regression equation. New predictors that add validity beyond that already afforded by AFQT are more likely to prove useful for selection purposes.

The analyses involved two steps. First, the criterion variable was regressed on AFQT. Second, we added all of the scales constituting the experimental predictors (e.g., RBI, Work Values) to the regression equation in the same step. The difference between the correlation (r) produced by the AFQT-only model and the multiple correlation (R) produced by the AFQT-plus-predictors model was then calculated to determine incremental validity.

We used the change in the log-likelihood ratio test to assess whether the OLS incremental validity coefficients (ΔR) were statistically significant. Traditional regression approaches rely on the F -test associated with the change in R^2 to assess significance. However, the computation of the R^2 change F -test depends on the sample size, such that large sample sizes tend to lead to more significant results. When using FIML missing data analyses, no single value of N applies to the entire sample. Using the full sample size (including cases with incomplete data) or relying on the listwise sample size is likely to result in Type I and Type II errors, respectively. As a result, when using FIML, researchers recommend relying on the Log-likelihood to assess how well each model fits the data (Enders, 2001). The likelihood value is an assessment of how well the estimated parameters fit the observed data. Multiplying the \log of the likelihood by -2 (-2 Log-likelihood or -2LL) yields a chi-square statistic that can be modeled using the standard assumptions underlying the chi-square distribution. We used the difference between the -2 Log-likelihood values from the AFQT-only model and the AFQT-plus-predictor model to determine whether the models significantly differed and thus whether the addition of the predictor set significantly improved model fit.

To make the interpretation of the results more comparable to previous SelectOCS projects (Russell et al., 2011; Russell & Tremble, 2011) we decided it would also be beneficial to present the significance of the F -test associated with the change in R^2 , despite the ambiguities associated with the sample sizes noted above. In an effort to be conservative, the listwise sample size (i.e., enlistment option = 203-682 and in-service = 160-196, depending on the predictor and criterion) was used for computing the F -test associated with change in R^2 . Although results of both the -2 Log-likelihood test and the F -test are presented, we relied more heavily on the more conservative -2 Log-likelihood test in our interpretation of model significance.

Range Restriction

As discussed in Chapter 2, the range on the AFQT is restricted for the current sample. Range restriction on AFQT, uncorrected, will lead to underestimates of the validity of AFQT and likely overestimates of incremental validity. Therefore, we corrected the regression coefficient between the candidates' AFQT score and each criterion for direct range restriction using Lawley's multivariate range restriction formula (Hunter & Schmidt, 1990; Lawley, 1943). Similar to analyses conducted in SelectOCS Phase 1 and Phase 2, the population AFQT standard deviation estimates were derived from the 1997 National Longitudinal Survey of Youth (NLSY97) (Bureau of Labor Statistics, 2005), a research project commissioned by the Department of Labor that includes data for 8,984 youths on a variety of variables, including college experience and ASVAB subtest scores. We used the college experience samples to derive the population standard deviation estimates for our samples. Specifically, we used youths from the NLSY97 data who were "college graduates" (16 or more years of schooling) as the population sample for the enlistment option candidates and those with "some college" as the population sample for the in-service candidates in our sample.

We completed the linear regression analyses incorporating the range restriction correction by creating two correlation matrices – one for the in-service sample and one for the enlistment option sample. The correlation matrices were estimated using FIML missing data analyses. We corrected all of the AFQT intercorrelations for direct range restriction. We then performed OLS

regression analyses on the corrected correlation matrices. The change in multiple R from step one to step two served as the primary diagnostic to evaluate each model.

Shrinkage

Sample-specific error could potentially inflate the estimates of R for predictor measures with small sample sizes and many scales. As a result, variations in sample sizes and the number of scales constituting each predictor measure make cross-measure comparisons difficult. To address this issue, we adjusted the observed incremental validity estimates using Burket's (1964) formula for shrinkage (cf. Formula 8; Schmitt & Ployhart, 1999). Calculating the corrected incremental validity estimates involved two additional steps:

1. Using the corrected correlations among the experimental predictor, AFQT, and the selected criteria, adjust the correlations between the predictors and continuously scaled criteria for sample size and number of predictors using Burket's (1964) formula for shrinkage.
2. Calculate the corrected incremental validity estimates for the experimental predictors by subtracting the shrunken R (the R from Step 1) associated with an AFQT-only model from the shrunken R obtained from the AFQT plus predictor model.

In subsequent analyses, the incremental validity coefficients corrected for direct range restriction on AFQT and shrinkage constitute the more conservative "corrected" coefficients, while those not corrected constitute the "uncorrected" coefficients.

Results

Cross-validation of Phase 2 and Phase 3 Results

Cohort Differences

Table 3.1 presents the mean differences between the Phase 2 and Phase 3 cohorts on the predictor and criterion measures. There were a number of statistically significant mean differences on the RBI, with the Phase 3 sample providing higher scores than Phase 2 sample on the Achievement, Fitness Motivation, Self-Efficacy, Affective Commitment, Interest in Leadership, Leader Self-Efficacy, and Tolerance for Injury scales. The Phase 2 sample provided higher scores than Phase 3 on the RBI Equity Sensitivity scale. Additionally, the Phase 3 sample had higher means on the Benevolence and Social Work Environment Values scales, both of which were presented with the RBI scales in the OBEF. There were also a number of statistically significant differences between the Phase 2 and Phase 3 samples on the criterion measures. The Phase 3 sample tended to indicate longer Career Intentions, higher Affective Commitment, higher Continuance Commitment, and higher Morale. On the other hand, the Phase 2 sample tended to rate themselves as having higher Satisfaction with their Branch Assignment.

Table 3.1. *Cohort Differences between the Phase 2 and Phase 3 Enlistment Option Cohorts on OBEF Scales*

Scale/Predictor/Criterion	Phase 2 (P2)		Phase 3 (P3)		P3-P2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d^f</i>
AFQT	88.44	9.02	88.79	8.39	.04
RBI Scales					
Peer Leadership	3.96	.53	4.02	.53	.10
Achievement	4.16	.48	4.23	.45	.16
Fitness Motivation	3.80	.63	3.94	.58	.22
Hostility to Authority	1.84	.51	1.89	.49	.11
Generalized Self-Efficacy	4.27	.52	4.34	.47	.15
Affective Commitment	3.87	.63	4.02	.59	.23
Stress Tolerance	3.26	.54	3.28	.53	.04
ARC Hostility to Authority	2.78	.62	2.84	.63	.10
Interest in Leadership	3.81	.71	3.95	.71	.20
Leadership Self-Efficacy	4.02	.43	4.11	.45	.19
Equity Sensitivity	2.46	.60	2.30	.57	-.26
Tolerance for Injury	3.64	.65	3.82	.61	.27
Work Values					
Benevolence (Likert)	4.08	.55	4.23	.55	.26
Social Work Environment (Likert)	3.71	.62	3.83	.60	.20
Selfless Service (Rank)	.70	1.21	.80	1.18	.08
Leadership (Rank)	.26	1.12	.23	1.17	-.02
Recognition (Rank)	-.04	1.09	-.01	1.08	.03
Pay (Rank)	.12	1.02	.07	1.03	-.05
Structure (Rank)	-.05	1.04	-.18	1.06	-.12
Comfort (Rank)	-.16	1.06	-.28	1.05	-.11
Home (Rank)	-.17	1.02	-.12	.99	.05
Challenge (Rank)	.00	.98	.05	.98	.05
Self-Direction (Rank)	-.09	.98	-.11	.91	-.01
Teamwork (Rank)	-.09	.87	-.15	.92	-.07
Variety (Rank)	.11	.88	.11	.87	.00
LKT					
Skills	9.39	.13	9.39	.14	.01
Traits	9.46	.16	9.45	.16	-.07
Army Identity Structure					
Overlap	5.25	1.19	5.40	1.12	.13
Self-Concept	4.06	1.57	4.28	1.48	.14
Conflict	4.92	1.32	5.04	1.40	.10
TAPAS					
Even Tempered	.15	.56	.12	.59	-.07
Curiosity/Continuous Learning	.08	.71	-.03	.70	-.16
Tolerance	-.57	.82	-.61	.86	-.05
Trust/Cooperation	-.50	.57	-.56	.55	-.10
Optimism	.10	.68	.23	.70	.20
Adjustment	-.10	.71	-.05	.68	.07
Dominance/Leadership	.00	.65	.07	.59	.10
Physical Condition	.47	.82	.52	.76	.06
Achievement	.29	.62	.34	.62	.07
Non-Delinquency	-.11	.62	-.15	.59	-.07
Responsibility	-.15	.61	-.08	.60	.12
Intellectual Efficiency	.25	.73	.12	.73	-.18

Table 3.1. (*Continued*)

Scale/Predictor/Criterion	Phase 2		Phase 3		P3-P2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i> ¹
Criterion Measures					
Career Intentions	2.88	.90	2.96	.88	.09
Affective Commitment	3.74	.78	3.97	.68	.29
Continuance Commitment	2.83	1.02	3.04	.92	.21
Morale	4.04	.75	4.23	.69	.26
Branch Satisfaction	4.61	.79	4.42	.99	-.24
Peer Ratings of Leadership Potential	5.42	1.33	5.65	1.37	.17
OCS Total Score (Order of Merit) ²	-.01	1.00	-.11	.95	-.10
OCS Leadership Score ²	-.05	.95	-.12	.94	-.07
OCS Physical Fitness ²	-.03	1.05	-.04	1.01	-.01
OCS Academic Performance ²	.18	.90	-.05	1.03	-.25
OCS APFT Score	263.08	23.13	276.65	18.32	.59
OCS Average History Score	87.25	7.93	85.57	8.95	-.21

Note. $d = (M_{\text{COMPARISON}} - M_{\text{REFERENT}}) / SD_{\text{REFERENT}}$, where Phase 2 is the referent group. Results are based on the raw data for Phase 2 and Phase 3 samples (enlistment option candidates only). Phase 2 $n = 429$, Phase 3 $n = 319$.

¹ Bolded values indicate statistical significance at $p < .05$. Significance is based on the independent samples *t*-test difference between the two means.

² Scores were standardized within OCS class

Because the OCS scores are standardized within class to remove any variation associated with classes, interpretation of the Phase 2 and Phase 3 OML comparisons is problematic. Each class is standardized to have a mean of zero and a standard deviation of one. If an average student in one class is higher performing than an average student in another class, the difference is removed after standardizing because both “average” students’ standardized scores will be zero. Direct comparisons among classes or phases will not be meaningful because most of the raw data are generally not standardized (i.e., each company/class has some discretion in how certain events are scored and combined to create the final OML scores). Therefore, the relations between predictor measures and the OCS standardized scores will not capture true differences between the two samples, which could result in erroneous differences when examining the differential validity estimates between the two samples. This is particularly true if there are policy differences between Phase 2 and Phase 3 that would lead one set to be systematically different from another set, as these differences would be masked by the standardization process. Evidence for this is seen in Table 3.1, which shows significant Phase 2 and Phase 3 differences on non-standardized attitudinal and performance outcomes (e.g., Affective Commitment, APFT), but not the outcomes that were standardized within class.

For this reason, we used two additional variables as surrogates of the OCS scores (see Chapter 2 for more details). First, we constructed an average APFT score – the mean of the initial APFT score and final APFT score. We used this as a surrogate for the OCS Physical Fitness score. As a surrogate for OCS Academic Performance, we took the average score between the two History courses. We chose history because those courses tend to have more variance in the scores among candidates and should have more consistent scoring because the same teacher has taught the course for several years. Peer Ratings of Leadership Potential were used a surrogate for OCS Leadership Performance scores. The correlations between the OCS measures of performance and the surrogates was high, ranging from .51 (i.e., the correlation between Peer Ratings of Leadership Potential and OCS Leadership in the Phase 3 sample) to .96

(i.e., the correlation between average APFT and OCS Physical Fitness in the Phase 2 sample) with an average of .75 (see Appendix C for complete results).

Multi-Group Measurement Equivalence

Table 3.2 lists the results of the multi-group measurement equivalence analyses.¹³ A chi-square difference test was used to examine the difference in regression parameter estimates when the parameters were unconstrained (i.e., left to be freely estimated; Not Equal) and when they were constrained to be equal between the Phase 2 and Phase 3 samples (Equal). Among the predictor sets, the only models to show significant differences between Phase 2 and Phase 3 were the regression of average APFT score on RBI and the regressions of the OML scores on RBI. This suggests that the validity coefficients differ between the two samples. As discussed previously, the standardization of the OML scores within class in combination with the differences in RBI scores may be contributing to the difference in the validity results. The non-significant differences between the two models for the remainder of the criteria and predictors sets suggest there was no drop in fit when the parameters were constrained to be equal. Based on parsimony, the model with the lower number of parameter estimates is considered better fitting, which is the model with equal parameter estimates. That is, with the exception of RBI and APFT score, the predictive relations between each set of predictors and the criteria are similar for the Phase 2 and Phase 3 samples..

We examined the non-equivalent validity coefficients in more detail for the relation between APFT and RBI, and found the relation stronger in Phase 3. When the model parameters are left to be freely estimated, the R^2 value for Phase 2 is .355 and .338 for Phase 3. When the regression parameters are constrained to be equal, the R^2 value for Phase 2 is .288 and .329 for Phase 3. Overall, the R^2 values for both models are statistically significant and large. The primary objective of cross-validation is to show that the predictive utility of a predictor set in one sample does not change in an independent sample. If this were the case, we would worry that the significance in the first sample may have been due to chance.

Among the RBI scales, Peer Leadership ($\Delta r = .15$), Hostility to Authority ($\Delta r = -.16$) Generalized Self-Efficacy ($\Delta r = .09$), Interest in Leadership ($\Delta r = .14$) and Leader Self-Efficacy ($\Delta r = .16$) showed stronger correlations with key outcomes in Phase 3 than Phase 2. Achievement correlated significantly with APFT in Phase 2 but not in Phase 3 ($\Delta r = -.14$). With the exception of Achievement, the relation between RBI and APFT increased in Phase 3. Overall, given only one difference on one predictor between the Phase 2 and Phase 3 samples, we feel confident combining the two samples to proceed with examining the incremental validity of the predictor sets. The correlations for Phase 2 and Phase 3 for all of the predictor sets and criteria can be found in Appendix C.

¹³ Because these analyses involve scale-dependent models (Cudeck, 1989), all analyses used variance/covariance matrices.

Table 3.2. *Measurement Invariance Cross-Validation Analyses of Phase 2 and Phase 3 Enlistment-Option Cohort*

	Not Equal		Equal			
	χ^2	df	χ^2	df	$\Delta \chi^2$	Δdf
AFQT + RBI						
Career Intentions	14,844.28	2,470	14,852.85	2,483	8.57	13
Affective Commitment	14,803.68	2,470	14,815.43	2,483	11.75	13
Leadership Potential	14,823.05	2,470	14,841.42	2,483	18.37	13
OCS History Score	14,775.37	2,470	14,795.84	2,483	20.47	13
OCS Academic Performance	14,721.14	2,470	14,745.34	2,483	24.20*	13
OCS Fitness Performance	14,634.92	2,470	14,667.96	2,483	33.04*	13
OCS Leadership Performance	14,783.90	2,470	14,813.24	2,483	29.34*	13
OCS APFT Score	14,707.83	2,470	14,738.83	2,483	31.00*	13
AFQT + TAPAS						
Career Intentions	16,790.13	2,470	16,802.86	2,483	12.73	13
Affective Commitment	16,791.05	2,470	16,803.39	2,483	12.34	13
Leadership Potential	16,762.87	2,470	16,776.84	2,483	13.97	13
OCS History Score	16,705.71	2,470	16,715.32	2,483	9.61	13
OCS Academic Performance	16,658.41	2,470	16,672.17	2,483	13.77	13
OCS Fitness Performance	16,648.96	2,470	16,661.70	2,483	12.74	13
OCS Leadership Performance	16,730.77	2,470	16,749.16	2,483	18.39	13
OCS APFT Score	16,686.08	2,470	16,700.27	2,483	14.19	13
AFQT + Work Values						
Career Intentions	15,670.09	2,442	15,677.74	2,456	7.64	14
Affective Commitment	15,651.55	2,442	15,662.38	2,456	10.83	14
Leadership Potential	15,656.63	2,442	15,670.45	2,456	13.81	14
OCS History Score	15,572.85	2,442	15,581.87	2,456	9.02	14
OCS Academic Performance	15,519.15	2,442	15,537.45	2,456	18.30	14
OCS Fitness Performance	15,676.29	2,442	15,683.90	2,456	7.61	14
OCS Leadership Performance	15,650.17	2,442	15,665.40	2,456	15.23	14
OCS APFT Score	15,677.85	2,442	15,689.47	2,456	11.62	14
AFQT + LKT						
Career Intentions	17,424.42	2,640	17,427.11	2,643	2.69	3
Affective Commitment	17,431.48	2,640	17,432.83	2,643	1.35	3
Leadership Potential	17,421.72	2,640	17,422.96	2,643	1.24	3
OCS History Score	17,323.18	2,640	17,325.05	2,643	1.87	3
OCS Academic Performance	17,281.06	2,640	17,283.95	2,643	2.89	3
OCS Fitness Performance	17,438.81	2,640	17,441.79	2,643	2.98	3
OCS Leadership Performance	17,428.87	2,640	17,431.23	2,643	2.36	3
OCS APFT Score	17,438.88	2,640	17,441.74	2,643	2.86	3
AFQT + Army Identity Structure						
Career Intentions	17,174.54	2,632	17,180.17	2,636	5.63	4
Affective Commitment	17,178.79	2,632	17,182.76	2,636	3.97	4
Leadership Potential	17,237.82	2,632	17,241.31	2,636	3.49	4
OCS History Score	17,140.11	2,632	17,142.44	2,636	2.33	4
OCS Academic Performance	17,089.89	2,632	17,098.00	2,636	8.12	4
OCS Fitness Performance	17,257.48	2,632	17,260.13	2,636	2.64	4
OCS Leadership Performance	17,247.00	2,632	17,250.51	2,636	3.52	4
OCS APFT Score	17,258.39	2,632	17,259.91	2,636	1.52	4

Note. $\Delta\chi^2$ = the difference in the chi-square model fit index when the regression parameters are unconstrained for the Phase 2 and Phase 3 samples (Not Equal) and when the regression parameters are constrained to be equal for Phase 2 and Phase 3 samples (Equal). *Significant values indicate that the constrained model exhibited significantly worse fit than did the unconstrained model, suggesting the validity coefficients are not similar for the two groups.

Incremental Validity

The incremental validity results are organized by service option (i.e., in-service and enlistment option) and predictor set. RBI, TAPAS, Work Values, LKT and Army Identity Structure were examined as predictors of candidates' Career Intentions, Affective Commitment, Peer Ratings of Leadership Potential, OCS APFT scores, and OCS History scores. First, we discuss the incremental validity results. Second, we discuss the impact of correcting for range restriction and shrinkage on the validity coefficients. Finally, we discuss the specific scales that predict each criterion. Only significant results will be discussed in the text.

Enlistment Option Results

Table 3.3 reports the incremental validity results for the enlistment option sample. Appendix C presents the bivariate correlations for Phase 2 and Phase 3, and Appendix D presents the bivariate correlations produced for the in-service and enlistment option samples. AFQT was a significant predictor of Peer Ratings of Leadership Potential ($R = .15$) and OCS History Score ($R = .44$), accounting for 2% and 19% of the variance, respectively.

Rational Biodata Inventory. The RBI scales provided incremental prediction of Career Intentions ($\Delta R = .26$), Affective Commitment ($\Delta R = .45$), Peer Leadership ratings ($\Delta R = .24$), and OCS APFT Score ($\Delta R = .56$), accounting for an additional 6% to 31% of the variance in the criteria. After correcting for range restriction, there was a drop in the incremental utility of the RBI among the criteria by between two to six percentage points. However, the corrected ΔR values remained relatively large, and the RBI accounted for between 3% and 25% of the variance in the criteria.

Many RBI scales related to the performance criteria and the continuance criteria (see Appendix D). Goal Expectations (mean $r = .19$), Affective Commitment (mean $r = .37$), and Generalized Self-Efficacy (mean $r = .15$) showed the strongest correlations with Career Intentions and with Affective Commitment. Fitness Motivation (mean $r = .41$), Tolerance for Injury (mean $r = .22$), Emotional Stability (mean $r = .15$) and Goal Expectations (mean $r = .17$) showed the strongest correlations with Peer Ratings of Leadership Potential and with APFT.

TAPAS. The TAPAS scales provided incremental prediction of Career Intentions ($\Delta R = .11$), Commitment ($\Delta R = .23$), Peer Leadership ratings ($\Delta R = .15$), and OCS APFT Score ($\Delta R = .49$), accounting for an additional 6% to 31% of the variance in the criteria. There was a slight drop in incremental prediction once corrected for range restriction and shrinkage. Among all of the criteria, the variance accounted for dropped by only one percent.

Several TAPAS scales related significantly to the criteria. Non-Delinquency showed the strongest relation with Career Intentions ($r = .13$) and Affective Commitment ($r = .16$). Physical Fitness showed the strongest relation with Peer Ratings of Leadership Potential ($r = .16$) and APFT ($r = .44$). Achievement ($r = .14$) and Responsibility ($r = .13$) showed modest relations with Affective Commitment. Achievement ($r = .15$) also showed modest relations with APFT.

Table 3.3. *Incremental Validity Results for the Enlistment Option Sample with Phase 2 and Phase 3 Samples Combined*

	ENLISTMENT (uncorrected)			ENLISTMENT (corrected)		
	AFQT	AFQT + predictor	ΔR	AFQT	AFQT + predictor	ΔR
<i>RBI</i>	.16	.47	.31	.23	.49	.26
Career Intentions	.14	.40	.26*	.23	.39	.16
Affective Commitment	.03	.48	.45*	.00	.42	.42
Leadership Potential	.15	.39	.24*	.23	.40	.16
OCS APFT Score	.03	.59	.56*	.07	.56	.50
OCS History Score	.44	.48	.05	.62	.68	.06
<i>TAPAS</i>	.16	.36	.20	.23	.41	.18
Career Intentions	.14	.26	.11*	.23	.27	.04
Affective Commitment	.03	.26	.23*	.00	.20	.20
Leadership Potential	.15	.30	.15*	.23	.41	.17
OCS APFT Score	.03	.52	.49*	.07	.56	.49
OCS History Score	.44	.45	.01	.62	.64	.02
<i>Work Values</i>	.16	.29	.13	.23	.30	.07
Career Intentions	.14	.24	.09	.23	.24	.01
Affective Commitment	.03	.30	.27*	.00	.25	.25
Leadership Potential	.15	.24	.09*	.23	.26	.03
OCS APFT Score	.03	.21	.18	.07	.13	.06
OCS History Score	.44	.46	.02	.62	.62	.00
<i>LKT</i>	.16	.18	.03	.23	.25	.03
Career Intentions	.14	.16	.02	.23	.25	.03
Affective Commitment	.03	.12	.09*	-.02	.09	.11
Leadership Potential	.15	.16	.01	.23	.23	.00
OCS APFT Score	.03	.04	.01	.07	.05	.00
OCS History Score	.44	.44	.00	.62	.64	.02
<i>Army Identity Structure</i>	.16	.28	.13	.23	.34	.11
Career Intentions	.14	.37	.23*	.23	.38	.15
Affective Commitment	.03	.38	.35*	.00	.38	.38
Leadership Potential	.15	.17	.02	.23	.24	.01
OCS APFT Score	.03	.06	.03	.07	.07	.01
OCS History Score	.44	.44	.00	.62	.62	.00

Note. Results are based on FIML regression results. Italicized values are the average multiple correlation coefficients for each predictor set across criteria. Bolded multiple correlations (i.e., R) indicate that the uncorrected version of these statistics were statistically significant at $p < .05$; Bolded ΔR are significant based on the F -test for R^2 change; *Significance based on the difference between the -2 Log-likelihood values for the AFQT-only model and the AFQT + predictor model. A significant value indicates that the AFQT + predictor model provided significantly better fit than did the AFQT-only model. $N = 748$.

Work Values. The Work Values scales provided incremental prediction of Commitment ($\Delta R = .27$) and Peer Ratings of Leadership Potential ($\Delta R = .09$), accounting for an additional 6% and 1% of the variance, respectively. After correcting the results for range restriction and shrinkage, the incremental prediction dropped by about one percentage point.

Benevolence, Social Work Environment, Teamwork, and Variety related significantly with both Affective Commitment (mean $r = .15$) and Peer Ratings of Leadership Potential (mean $r = .09$). In addition, Selfless-Service ($r = .17$), Home ($r = .09$), and Challenge ($r = .13$) related significantly to Affective Commitment.

Leadership Knowledge Test. The LKT provided incremental prediction of Affective Commitment ($\Delta R = .09$) only, accounting for an additional 1% of the variance. There was no drop in incremental prediction after correcting the relation for range restriction and shrinkage. The LKT Traits scale showed the strongest relation with of Affective Commitment ($r = .10$).

Army Identity Structure. Army Identity Structure provided incremental prediction of Career Intentions ($\Delta R = .23$) and Affective Commitment ($\Delta R = .35$), accounting for 5% and 12% of the variance, respectively. Incremental validity of Career Intentions dropped three percentage points after correcting the relation for range restriction and shrinkage, but overall the change in R remained notable ($\Delta R = .15$).

All three Army Identity Structure scales (Concept, Conflict, and Overlap) related significantly to both Career Intentions and Affective Commitment. Overlap ($r = .34$) showed the strongest relations with Career Intentions, and Conflict ($r = .34$) showed the strongest relation with Affective Commitment.

Summary. The RBI and TAPAS emerged as the strongest predictors of the criteria. Both measures provided incremental validity to the prediction of all criteria except the OCS History score. However, the RBI generally afforded more predictive efficacy (mean $\Delta R = .31$) than the TAPAS (mean $\Delta R = .20$). This is at least in part because the RBI's Affective Commitment scale is very similar to the Affective Commitment and Career Intentions criterion variables. Those criteria aside, the RBI has a slight advantage over the TAPAs for predicting other criteria. Work Values, LKT, and Army Identification each provided incremental prediction of Affective Commitment but afforded little additional value for any of the performance criteria (Peer Ratings of Leadership Potential, APFT, or History Score). Army Identity Structure provided incremental prediction of the continuance criteria, suggesting that the Army Identity Structure scales could help identify individuals who are likely to stay with and be more committed to the Army. Although there was a decrease in incremental prediction once the correlations were corrected for range restriction and the R -values were corrected for shrinkage, the overall findings did not change.

In-Service Results

Table 3.4 reports the incremental validity results for the in-service sample. Appendix D presents the bivariate correlations produced from the FIML analyses. AFQT was a significant predictor of OCS History Score ($R = .62$), accounting for 37% of the variance.

Rational Biodata Inventory. The RBI scales provided incremental prediction of Career Intentions ($\Delta R = .23$), Affective Commitment, ($\Delta R = .47$), Peer Ratings of Leadership Potential ($\Delta R = .23$), and OCS APFT score ($\Delta R = .23$), accounting for an additional 5% and 22% of the variance. After correcting for range restriction and shrinkage, the incremental prediction of Peer Leadership and APFT score dropped from .23 to .12, thus accounting for only 1% of the variance in the criterion—a decrease of 4 percentage points. Similarly, the incremental prediction of Affective Commitment decreased by eight percentage points.

With the exception of Peer Leadership and Fitness Motivation, all RBI scales correlated significantly with Career Intentions in the theoretically expected direction. With the exception of Interest in Leadership, all of the RBI scales also related significantly to Affective Commitment in the theoretically expected direction. Hostility to Authority (mean $r = -.19$), ARC Hostility to Authority (mean $r = -.20$), and Equity Sensitivity (mean $r = -.28$) related negatively to Career Intentions and Affective Commitment. Other scales that related strongly to the continuance criteria include Affective Commitment (mean $r = .31$), Generalized Self-Efficacy (mean $r = .22$), and Stress Tolerance (mean $r = .19$). Fitness Motivation (mean $r = .19$) and Tolerance for Injury (mean $r = .15$) showed the strongest relations with the Peer Ratings of Leadership Potential and APFT.

TAPAS. The TAPAS scales provided incremental prediction of Peer Ratings of Leadership Potential ($\Delta R = .24$) and APFT ($\Delta R = .28$), accounting for an additional 6% and 8% of the variance in the criteria. The F-test indicated that the TAPAS provided incremental prediction of Career Intentions, however the -2 Log-likelihood indicate the model was not significant ($\Delta R = .24$). The incremental prediction of all three criteria dropped after correcting for range restriction.

Fewer TAPAS scales related to the criteria for the in-service sample than in the enlistment option sample. Non-Delinquency related negatively to Peer Ratings of Leadership Potential ($r = -.14$), and Responsibility related positively to Peer Ratings of Leadership Potential ($r = .19$). Physical Fitness was the only TAPAS scale to emerge as a significant predictor of APFT.

Work Values. The Work Values scales provided incremental prediction of Career Intentions ($\Delta R = .23$), Affective Commitment ($\Delta R = .36$), and Peer Ratings of Leadership Potential ($\Delta R = .17$). There was a slight drop in incremental prediction after correcting for range restriction and shrinkage. Most notably the incremental prediction of Affective Commitment dropped by seven percentage points, but the ΔR remained notable at .23.

Table 3.4. *Incremental Validity Results for the In-Service Sample with Phase 2 and Phase 3 Samples Combined*

	IN-SERVICE (uncorrected)			IN-SERVICE (corrected)		
	AFQT	AFQT + predictor	ΔR	AFQT	AFQT + predictor	ΔR
<i>RBI</i>	.24	.48	.24	.25	.42	.17
Career Intentions	.15	.38	.23*	.06	.27	.21
Affective Commitment	.05	.52	.47*	.07	.44	.37
Leadership Potential	.21	.44	.23*	.23	.35	.12
OCS APFT Score	.20	.43	.23*	.22	.34	.12
OCS History Score	.61	.65	.04	.66	.68	.02
<i>TAPAS</i>	.25	.45	.20	.25	.32	.08
Career Intentions	.15	.39	.24	.06	.19	.13
Affective Commitment	.05	.24	.18	.07	.00	.00
Leadership Potential	.23	.47	.24*	.23	.40	.16
OCS APFT Score	.21	.49	.28*	.22	.33	.11
OCS History Score	.62	.66	.04	.66	.66	.00
<i>Work Values</i>	.25	.44	.19	.25	.35	.10
Career Intentions	.15	.38	.23*	.06	.26	.20
Affective Commitment	.04	.40	.36*	.07	.31	.23
Leadership Potential	.23	.40	.17*	.23	.29	.06
OCS APFT Score	.21	.35	.14	.22	.22	.00
OCS History Score	.62	.66	.04	.66	.67	.00
<i>LKT</i>	.25	.31	.06	.25	.29	.05
Career Intentions	.15	.19	.04	.06	.10	.04
Affective Commitment	.05	.30	.25*	.07	.28	.21
Leadership Potential	.23	.23	.01	.23	.21	.00
OCS APFT Score	.20	.21	.01	.22	.19	.00
OCS History Score	.62	.62	.00	.66	.66	.00
<i>Army Identity Structure</i>	.25	.38	.14	.25	.37	.12
Career Intentions	.16	.28	.12*	.06	.25	.19
Affective Commitment	.03	.38	.35*	.07	.37	.30
Leadership Potential	.23	.30	.08*	.23	.28	.05
OCS APFT Score	.19	.32	.13*	.22	.30	.08
OCS History Score	.62	.63	.01	.66	.66	.00

Note. Results are based on FIML regression results. Bolded multiple correlations (i.e., R) indicate that the uncorrected version of these statistics were statistically significant at $p < .05$; Bolded ΔR are significant based on the F -test for R^2 change; *Significance based on the difference between the -2 Log-likelihood values for the AFQT-only model and the AFQT + predictor model. A significant value indicates that the AFQT + predictor model provided significantly better fit than did the AFQT-only model. Italicized values are the average multiple correlation coefficients for each predictor set across criteria. $N = 241$.

Pay related negatively to Career Intentions ($r = -.20$), whereas Variety related positively to Career Intentions ($r = .19$). Benevolence ($r = .28$), Social Work Environment ($r = .14$), Selfless Service ($r = .13$), Teamwork ($r = .13$), and Variety ($r = .13$) all related significantly to Affective Commitment. Finally, Benevolence ($r = .14$) and Structure ($r = -.15$) were the driving predictor scales of Peer Ratings of Leadership Performance.

Leadership Knowledge Test. The LKT provided incremental prediction of Affective Commitment ($\Delta R = .25$). There was a slight drop in incremental validity after correcting for range restriction and shrinkage, but the ΔR value remained notable ($\Delta R = .21$). The LKT Traits scale was the strongest predictor of Affective Commitment ($r = .17$).

Army Identity Structure. The Army Identity Structure scales provided incremental prediction of Career Intentions ($\Delta R = .12$), Affective Commitment ($\Delta R = .35$), Peer Ratings of Leadership Potential ($\Delta R = .08$), and APFT ($\Delta R = .13$), accounting for between 1% and 12% of the variance. Incremental prediction dropped between zero and three percentage points across criteria after correcting for range restriction and shrinkage.

All three Army Identity Structure scales (Overlap, Conflict, and Concept) related significantly to the continuance criteria, with Overlap ($r = .31$) showing the strongest relation with Affective Commitment and Conflict showing the strongest relation with Career Intentions. The three scales showed weaker relations with the performance criteria, with Concept having the highest predictive efficacy (mean $r = .10$).

Summary. The RBI and Army Identity Structure scales provided incremental prediction of both continuance and performance criteria. However, the RBI generally afforded more predictive value (mean $\Delta R = .24$) than did the measures of Army Identity Structure (mean $\Delta R = .14$). Unlike the enlistment option sample, the TAPAS did not afford any additional variance to the prediction of the continuance criteria for those entering OCS through the in-service option. Similar to the enlistment option sample, the Work Values scales and the LKT each significantly improved the prediction of Affective Commitment but afforded little additional value for any of the performance criteria (Peer Ratings of Leadership Potential, APFT, or History Score). Overall, the RBI demonstrated the most predictive efficacy beyond AFQT for both continuance and performance criteria in both samples.

Summary and Conclusions

Two objectives of this research were to (a) cross-validate the SelectOCS Phase 2 results with Phase 3 data and (b) select the most promising individual instrument for predicting officer performance and continuance. This chapter empirically addressed both objectives.

With regard to the first objective, the multi-group measurement equivalence analyses suggest that the OBEF validity estimates were quite consistent from Phase 2 to Phase 3. The exception to this was the RBI, which had different results for multiple OCS performance measures. However, the nature of the subgroup difference and multi-group results suggest that the differences on the standardized-within-class criteria could be attributed to differences in the validity estimates. Of the criteria that were not standardized, the only difference was for APFT. For that criterion, the validity coefficients in Phase 3 were generally in the same direction but larger. This suggests that overall, the OBEF validity estimates were consistent from Phase 2 to Phase 3 and subsequent analyses should combine these two samples.

With regard to the second objective, the most promising individual measure for predicting performance and continuance was the RBI. RBI scales predicted both continuance and performance criteria as well as or better than other instruments, even when controlling for shrinkage. Furthermore, these results held for both the enlistment option and in-service samples. There are a number of considerations in selecting an instrument for operational use, including (a) criterion-related validity, (b) administration time, (c) adverse impact, and (d) susceptibility to faking and coaching. The validity analyses presented here suggest that, based solely on the criterion-related validity of each component measure, the RBI should be the primary instrument tested for operational use, because it holds the most promise for predicting key outcomes of interest. Final decisions on what instrument to administer operationally should consider the other three factors as well. Although the RBI should be the primary instrument, aspects of other OBEF instruments (e.g., TAPAS, Work Values) might contribute incrementally to a composite developed using only RBI scales. We explore this possibility in the next chapter.

Chapter 4: Development and Evaluation of Revised OBEF Composites

This chapter builds on the validity analyses conducted in Chapter 3 by determining which set of OBEF measures best balances both prediction of key criteria and parsimony. This satisfies the third objective described in Chapter 1. Given the small sample size, we could not conduct composite formation analyses for the in-service sample; therefore, the results presented apply to the enlistment option sample only.

Approach

Among the RBI, TAPAS, Work Values, LKT, and Army Identification measures, 47 scales could be included in an empirical selection composite. The large number of possible scales presents several challenges both empirically and practically. First, there is likely to be a fair amount of overlap among many of the scales, although each is likely to account for only a small portion of variance in the criteria. Including all 47 scales into the composite formation analyses would result in a high likelihood that none of the scales emerge as driving predictors of the criteria. Furthermore, it is unlikely that the Army would administer all five OBEF measures operationally in the future, due to practical considerations such as testing time. Given these factors, one goal of these analyses was to identify a composite using only one measure and then determine what additional scales could be included if additional time was available to administer multiple measures or if new scales were to be developed. As a result, we used one predictor measure as a “base” for the selection composite. As described in Chapter 3, we identified the base measure by examining the incremental validity results and determining which measure provided the most predictive efficacy among both performance and continuance criteria. Based on the results in Chapter 3, we chose the RBI as the base measure. We developed the composite in two stages. First, we identified scales of the RBI that best predicted the criteria. Second, we examined the other measures to determine which scales provided incremental validity to the RBI scales.

To cross-validate the composites, we randomly split the sample into an analysis sample and a holdout sample. The analysis sample comprised two-thirds of the sample ($n = 501$) and the holdout sample comprised the remaining one-third ($n = 246$). We used the analysis sample to identify best-bet scales to include in a prediction composite. We used the holdout sample to confirm the results once the composite was constructed. We independently identified two composites, one for the performance criteria (i.e. Peer Ratings of Leadership Potential, APFT, and History Score) and one for the continuance criteria (i.e. Career Intentions and Affective Commitment).

Procedure

For the composite formation analyses, as in Chapter 3, we constructed a correlation matrix using FIML missing data analyses. We used the correlation matrix as the data input for the subsequent regression analyses. Final evaluation of the composites was performed using FIML regression analyses in MPLUS. The composites were developed with the steps described below.

1. Identify RBI scales with the most predictive efficacy using relative importance analyses

We considered several indices of relative importance in the scale selection analyses. First, we examined the bivariate correlations of each predictor-criterion relation and the standardized regression coefficient (β) associated with the individual RBI scales when each criterion was regressed on all of the RBI scales using OLS regression. Second, using the full model containing all RBI scales as a baseline, we examined the decrease in model R^2 for each criterion when removing each scale from the full model. Thus, for each RBI scale, we examined the effect size: $-\Delta R^2 = R^2_{\text{FULL}} - R^2_{\text{RESTRICTED}}$, where $R^2_{\text{RESTRICTED}}$ is the model R^2 obtained when the RBI scale in question is removed from the full model. Third, we used the procedure developed by Johnson (2000; see also Johnson & LeBreton, 2004) to compute relative weights (RW). When re-scaled to a proportion metric ranging from 0.0%-100.0% (as has been done here), estimates can be interpreted as the percentage of criterion variance accounted for (R^2) by each RBI scale. Finally, we used best subsets regression to compute the regressions of all possible combinations of the RBI scales. We counted the number of times a particular scale was included in one of the top 10 regression models (rank ordered using the computed Mallows' C_p statistic; Mallows, 1973) and used this to evaluate the relative importance of that scale.

We computed all four metrics for each criterion of interest. To select RBI scales for inclusion in the empirical composite, we ensured that the correlation between the RBI scale and the criterion was in the theoretical direction among all criteria. We used this information to construct two initial RBI-based composites, one focused on predicting performance, the other on predicting continuance.¹⁴

2. Identify scales that add predictive utility to RBI composite

We examined the remaining RBI scales, the TAPAS, the Work Values scales, the LKT, and the Army Identity Structure scales to determine which added predictive power to the RBI composites developed in Step 1. We used a set of three analyses to narrow down and select a final subset of predictors to include in the final composite. First, we examined the semi-partial correlations between each of the scales and the criteria after controlling for the relation between the RBI composite and each criterion. We chose an initial set of predictors to examine based on whether each scale's semi-partial correlation was in the theoretically correct direction and was statistically significant (i.e., whether the predictor still afforded predictive variance once the relation between the RBI composite and the criterion was removed).

Next, we examined the relative importance of the subset of predictors identified through semi-partial correlations using Johnson's relative weights analysis. Similar to the RBI composites, we narrowed the set of additional predictors by (a) ensuring the correlation between

¹⁴ We performed additional steps to maximize the predictive efficacy of the RBI composite developed in this step. However, because the Army could use these scales operationally at a later date, any information regarding (a) which specific scales were included in the composite and (b) how those scales were combined to create the empirical selection composites has been taken out of this report. Readers interested in more specific information should contact the report authors.

each predictor and criterion was in the theoretically expected direction among all criteria and then (b) using the average Johnson's relative weight.

Finally, we used theoretical stepwise regression analysis to remove any scales that did not provide incremental prediction of the criteria. Specifically, we entered the AFQT and the RBI composite into the model first. Then, we entered each scale identified via Johnson's relative weights into the model one at a time based on the rank order of relative importance. We examined the ΔR value for each scale. We removed any scale that did not provide incremental prediction of the criteria from consideration. We used this information to create final empirical selection composites.

3. Evaluate the final composites

We cross-validated the final composites constructed in Step 2 against the holdout sample. First, we examined the bivariate correlations between the composites and the criteria to ensure the relations were in the same direction and had the same relative magnitude for the holdout sample and the analyses sample. Second, we examined the incremental validity of the final composites for predicting the target criteria. The criterion variables were first regressed on the AFQT, and then the final composites were added to the model. We assessed incremental validity by the ΔR value when the final composites were added to the model. We examined the incremental validity estimates of the holdout sample in comparison to the analysis sample to ensure the final composites afforded similar validity and statistical significance to the prediction of the criteria. Finally, the composites were examined for mean differences between subgroups using the same procedures described in Chapter 2.

Results

RBI Composites

Tables 4.1 and 4.2 list the relative importance results for the continuance criteria and the performance criteria, respectively. For the continuance criteria, Affective Commitment ($\overline{RW} = 48.85\%$),¹⁵ Goal Expectations ($\overline{RW} = 12.55\%$), Tolerance for Injury ($\overline{RW} = 7.75\%$), and Tolerance for Ambiguity ($\overline{RW} = 5.75\%$) accounted for the most variance in Affective Commitment and Career Intentions.

Across all performance criteria, Fitness Motivation ($\overline{RW} = 36.20\%$), Tolerance for Injury ($\overline{RW} = 8.00\%$), Tolerance for Ambiguity ($\overline{RW} = 7.00\%$), Goal Expectations ($\overline{RW} = 6.53\%$) and Stress Tolerance ($\overline{RW} = 4.73\%$) were in the theoretically correct direction and accounted for the most variance in all three criteria. Emotional Stability also accounted for a similar proportion ($\overline{RW} = 4.93\%$) of the variance as Stress Tolerance, but when examining the incremental validity of the scales, Emotional Stability, on average, did not provide incremental prediction among the performance criteria. Using the results, we developed two RBI-based composites, one for predicting performance and one for predicting continuance.

¹⁵ \overline{RW} = Average relative weight.

Table 4.1. *Relative Importance of the RBI Scales for Predicting Select Continuance Criteria in the Analysis Sample*

RBI scale	EOC Affective Commitment					EOC Career Intentions				
	<i>r</i>	<i>B</i>	$-\Delta R$	RW	T10	<i>r</i>	β	$-\Delta R$	RW	T10
Achievement Orientation	.17	-.05	.00	4.90%	4	.07	.07	.00	1.40%	5
Affective Commitment	.38	.30	.10	49.20%	10	.35	.41	.06	48.50%	10
ARC Hostility to Authority	-.11	-.03	.00	2.10%	3	-.07	-.05	.00	1.30%	
Emotional Stability	.11	-.03	.00	2.00%	3	.06	.06	.00	0.40%	
Equity Sensitivity	-.19	.03	.01	6.60%	10	-.07	-.09	.00	1.10%	
Fitness Motivation	.11	-.08	.00	2.20%	4	.07	.07	.00	1.20%	6
Goal Expectations	.13	.20	.00	2.60%		.26	-.02	.02	22.50%	10
Hostility to Authority	-.11	.00	.00	1.30%		-.04	.01	.00	0.30%	
Interest in Leadership	.12	.05	.00	1.80%		.14	.04	.00	3.50%	
Leadership Self-Efficacy	.12	-.12	.00	1.30%		.07	-.03	.01	1.90%	10
Narcissism	-.04	-.03	.00	0.80%		.01	-.03	.00	0.40%	
Peer Leadership	.07	-.02	.01	1.60%	10	.07	-.17	.00	0.90%	
Generalized Self-Efficacy	.14	-.03	.00	1.40%		.12	-.05	.00	1.90%	
Social Acuity	.11	.08	.01	3.70%	10	.10	.17	.00	2.30%	4
Stress Tolerance	.04	.06	.02	2.80%	10	.06	-.20	.00	1.10%	1
Tolerance for Ambiguity	.21	-.02	.02	11.20%	10	.04	.19	.00	0.30%	
Tolerance for Injury	.04	.09	.03	4.60%	10	.20	-.20	.01	10.90%	9

Note. Bolded values indicate statistical significance at $p < .05$; r = the bivariate correlation from the FIML correlation matrix; β = standardized regression weight when the predictor is entered with all of the other predictors in the second step of an OLS regression analysis; $-\Delta R$ = the decrease in R when the scale is removed from the OLS regression analysis; RW = Johnson's relative importance weight; T10 = The number of times the scale is included in the top 10 best subsets regression models. The analysis sample includes a random sampling of approximately 2/3 of the original sample ($n = 501$).

Table 4.2. *Relative Importance of the RBI Scales for Predicting Select Performance Criteria in the Analysis Sample*

RBI	Peer Ratings of Leadership Potential					OCS APFT Score					OCS History Score				
	<i>R</i>	β	$-\Delta R$	RW	T10	<i>R</i>	β	$-\Delta R$	RW	T10	<i>r</i>	β	$-\Delta R$	RW	T10
Achievement Orientation	.04	.01	.00	0.80%		.16	.14	.01	4.00%	10	-.05	-.04	.00	1.20%	2
Affective Commitment	.09	.02	.00	1.60%	7	.13	-.09	.01	1.20%	9	-.07	-.01	.00	5.90%	
ARC Hostility to Authority	-.06	-.03	.00	1.50%		-.07	-.07	.00	0.80%	9	-.03	-.06	.00	2.60%	5
Emotional Stability	.15	.10	.00	7.30%	8	.05	-.14	.01	1.40%	10	-.03	-.08	.00	6.10%	
Equity Sensitivity	-.07	-.01	.00	1.10%		-.09	-.04	.00	0.70%	1	-.01	.00	.00	1.50%	
Fitness Motivation	.27	.22	.04	35.80%	10	.54	.57	.24	68.20%	10	.04	.10	.01	4.60%	10
Goal Expectations	.13	.08	.00	7.40%	8	.24	.22	.03	10.50%	10	-.07	-.03	.00	1.70%	
Hostility to Authority	.00	.08	.00	1.70%	3	-.09	.01	.00	0.60%		.08	.10	.01	8.20%	10
Interest in Leadership	.05	.00	.00	0.90%		.06	-.01	.00	0.60%		-.09	-.07	.00	6.90%	10
Leadership Self-Efficacy	.03	-.09	.00	2.20%	3	.05	-.06	.00	1.10%	3	-.12	-.02	.00	9.10%	
Narcissism	-.07	-.08	.00	3.90%	9	-.02	-.05	.00	0.40%	5	.02	.09	.01	2.30%	6
Peer Leadership	.02	-.13	.01	2.90%	10	.03	-.12	.01	1.30%	7	-.07	-.07	.00	4.60%	4
Generalized Self-Efficacy	.12	.03	.00	2.90%		.09	-.14	.01	1.80%	10	-.10	-.09	.00	10.90%	10
Social Acuity	.09	.12	.01	4.50%	10	.04	.01	.00	0.50%		.03	.16	.01	7.80%	10
Stress Tolerance	.14	.09	.00	7.60%		.07	.00	.00	0.70%		.05	.05	.00	5.90%	
Tolerance for Ambiguity	.00	-.17	.02	6.30%	10	.09	.12	.01	1.50%	10	.08	.16	.02	13.20%	10
Tolerance for Injury	.19	.07	.00	11.70%	10	.20	.03	.00	4.80%		.06	.10	.01	7.50%	10

Note. Bolded values indicate statistical significance at $p < .05$; r = the bivariate correlation from the FIML correlation matrix; β = standardized regression weight when the predictor is entered with all of the other predictors in the second step of an OLS regression analysis; $-\Delta R$ = the decrease in R when the scale is removed from the OLS regression analysis; RW = Johnson's relative importance weight; T10 = The number of times the scale is included in the top 10 best subsets regression models. The analysis sample includes a random sampling of approximately 2/3 of the original sample ($n = 501$).

A full listing of incremental validity result for the Continuance and Performance RBI composites can be found in Table 4.3. The RBI Continuance composite provided significant incremental validity beyond the AFQT for predicting Career Intentions ($\Delta R = .29$) and Affective Commitment ($\Delta R = .32$). The RBI Performance Composite added significant incremental validity beyond the AFQT for predicting Peer Leadership Ratings ($\Delta R = .21$), OCS APFT average ($\Delta R = .49$), OCS Total score ($\Delta R = .28$), OCS Leadership Score ($\Delta R = .26$), and OCS Fitness Score ($\Delta R = .52$).

Table 4.3. *Incremental Validity Results for the RBI Composites*

	AFQT	RBI Continuance Composite	ΔR
Career Intentions	.08	.37	.29
Commitment	.04	.37	.32
	AFQT	RBI Performance Composite	ΔR
Leadership Potential	.09	.31	.21
OCS APFT average	.04	.54	.49
OCS History average	.44	.45	.01
OCS Total Score	.18	.46	.28
OCS Leadership Score	.09	.36	.26
OCS Fitness Score	.00	.52	.52
OCS Academic Score	.47	.48	.01

Note. Bolded values indicate statistical significance at $p < .05$.

Full Composite

We computed semi-partial correlations on the remaining RBI scales, the TAPAS scales, the Work Values scales, the LKT, and the Army Identity Structure scales. The semi-partial correlations reported describe the relation between a particular scale and a criterion after the relation between the RBI composite and the criterion is removed. Tables 4.4 and 4.5 list the scales that had significant semi-partial correlations that were in the theoretically correct direction for at least one of the continuance or performance criteria. The bivariate correlations between these scales and the criteria are also presented.

We computed Johnson's relative weights on the subset of scales identified in Tables 4.4 and 4.5. For the continuance criteria, seven scales accounted for the majority of the variance in either Career Intentions or Affective Commitment. For the performance criteria, five scales accounted for the majority of the variance in at least one of the performance criteria.

Table 4.4. *Bivariate Correlations, Semi-Partial Correlations, and Relative Weights for the Continuance Composite*

	Career Intentions			Affective Commitment		
	<i>r</i>	Semi Partial <i>r</i>	RW	<i>r</i>	Semi Partial <i>r</i>	RW
AFQT	-.08	-.03	0.70%	-.04	.00	0.20%
RBI Commitment Composite	.37		9.70%	.37		29.90%
Army ID Structure: Concept	.28	.10	15.30%	.27	.09	7.40%
Army ID Structure: Conflict	.27	.07	23.60%	.31	.12	11.40%
Army ID Structure: Overlap	.36	.18	22.30%	.30	.10	8.60%
LKT: Traits	.11	.07	0.60%	.09	.05	0.70%
TAPAS: Dominance	.08	.00	2.90%	-.02	-.10	1.40%
TAPAS: Intellectual Efficiency	-.07	-.05	0.30%	-.10	-.08	1.60%
TAPAS: Non-Delinquency	.14	.12	0.40%	.17	.16	3.90%
TAPAS: Optimism	.08	.01	2.20%	.15	.09	5.70%
TAPAS: Tolerance	.04	.08	1.60%	.09	.13	2.40%
TAPAS: Trust	-.03	.01	2.10%	.11	.16	3.00%
Work Values: Benevolence	.15	.02	8.10%	.24	.11	7.00%
Work Values: Challenge	.06	.06	0.20%	.09	.09	1.20%
Work Values: Comfort	.05	.04	0.20%	.09	.08	1.70%
Work Values: Pay	.07	.10	0.60%	-.02	.00	1.40%
Work Values: Selfless Service	.07	.04	0.30%	.20	.17	5.70%
Work Values: Team	.06	.03	0.30%	.12	.09	0.90%
Work Values: Variety	.03	.00	0.70%	.12	.09	2.20%

Note. Bolded values indicate statistical significance at $p < .05$; r = the bivariate correlation from the FIML correlation matrix; semi-partial r = the portion of the variance in the criteria accounted by the predictor over and beyond the portion of variance accounted for by the RBI composite; RW = Johnson's relative importance weight. Analyses were conducted on the enlistment option sample ($n = 501$).

Table 4.5. *Bivariate Correlations, Semi-Partial Correlations, and Relative Weights for the Performance Composite*

	Peer Leadership Ratings			Average APFT Score			Average History Score		
	<i>r</i>	Semi-Partial <i>r</i>	RW	<i>r</i>	Semi-Partial <i>r</i>	RW	<i>r</i>	Semi-Partial <i>r</i>	RW
AFQT	.10	.13	11.30%	.05	.10	3.50%	.44	.45	77.50%
RBI Performance	.28		44.30%	.53		45.50%	.04		1.80%
LKT: Traits	.11	.11	5.10%	-.01	-.01	0.10%	.06	.06	0.90%
LKT: Skills	.09	.09	3.30%	-.01	-.02	0.10%	.10	.10	1.80%
TAPAS: Achievement	.10	.04	3.80%	.19	.08	4.60%	-.09	-.10	3.50%
TAPAS: Curiosity	-.07	-.07	5.10%	-.03	-.03	0.30%	.09	.09	1.40%
TAPAS: Intellectual Efficiency	-.06	-.04	3.40%	-.12	-.10	3.40%	.20	.21	9.10%
TAPAS: Physical Fitness	.19	.01	13.20%	.46	.16	29.80%	.01	-.02	0.70%
TAPAS: Trust	-.05	-.01	1.60%	-.16	-.07	3.60%	-.04	-.03	0.60%
Work Values: Home	.07	.06	3.20%	.17	.15	5.70%	-.05	-.05	0.80%
Work Values: Recognition	-.04	-.05	2.50%	.10	.08	1.50%	.00	-.01	0.10%
Work Values: Self Direction	.03	.02	0.70%	.08	.08	0.80%	.02	.02	0.60%
Work Values: Structure	-.03	.00	0.70%	-.04	.00	0.70%	-.06	-.05	1.10%
Work Values: Variety	.06	.04	2.00%	.07	.03	0.40%	-.01	-.02	0.10%
	OCS Total Score			OCS Fitness Score			OCS Leadership Score		
	<i>r</i>	Semi-Partial <i>r</i>	RW	<i>r</i>	Semi-Partial <i>r</i>	RW	<i>r</i>	Semi-Partial <i>r</i>	RW
AFQT	.19	.23	17.10%	.02	.07	1.60%	.09	.13	8.60%
RBI Performance	.40		35.80%	.51		46.50%	.33		33.90%
LKT: Traits	.11	.11	2.10%	.00	.00	0.10%	.12	.12	4.10%
LKT: Skills	.10	.10	2.10%	.02	.02	0.50%	.08	.08	1.80%
TAPAS: Achievement	.16	.08	4.80%	.18	.08	4.30%	.16	.10	7.30%
TAPAS: Curiosity	-.06	-.06	2.30%	-.03	-.03	0.20%	-.10	-.10	5.50%
TAPAS: Intellectual Efficiency	-.03	-.01	1.30%	-.09	-.06	1.50%	-.08	-.06	3.20%
TAPAS: Physical Fitness	.36	.13	25.80%	.46	.17	33.40%	.29	.10	21.90%
TAPAS: Trust	-.09	-.02	1.60%	-.13	-.05	2.70%	-.06	-.01	1.30%
Work Values: Home	.13	.11	4.80%	.16	.14	5.20%	.12	.11	6.00%
Work Values: Recognition	.04	.03	0.20%	.12	.10	2.40%	.02	.01	0.20%
Work Values: Self Direction	.01	.01	0.20%	.08	.08	0.70%	-.03	-.03	1.80%
Work Values: Structure	.01	.04	0.20%	-.02	.02	0.40%	.05	.08	1.60%
Work Values: Variety	.10	.07	1.80%	.09	.05	0.60%	.10	.08	2.90%

Note. Bolded values indicate statistical significance at $p < .05$; r = the bivariate correlation from the FIML correlation matrix; semi-partial r = the portion of the variance in the criteria accounted by the predictor over and beyond the portion of variance accounted for by the RBI composite; RW = Johnson's relative importance weight. Analyses were conducted on the enlistment option sample ($n = 501$).

We conducted regression analyses to examine the incremental validity of each of the scales identified above. A theoretical stepwise regression approach was used where first AFQT and the RBI composite were entered into the model and then each scale was entered into the model one at a time in rank order starting with the predictor with the highest relative weight. We removed scales that did not provide incremental prediction to the criteria from the composite.

Finally, we conducted incremental validity analyses by first regressing the criteria on the AFQT and then adding the full composite. The full continuance composite provided significant incremental prediction to Career Intentions ($\Delta R = .33$) and Affective Commitment ($\Delta R = .38$). The full Performance Composite provided incremental prediction of Peer Ratings of Leadership Potential ($\Delta R = .23$), OCS APFT ($\Delta R = .52$), OCS Total score ($\Delta R = .31$), OCS Leadership Score ($\Delta R = .28$), and OCS Fitness Score ($\Delta R = .56$). A complete list of validity coefficients are provided in Tables 4.6 and 4.7.

Table 4.6. *Incremental Validity Results for the Full Continuance Composite*

Continuance Criteria	Analysis Sample				Holdout Sample			
	<i>r</i>	<i>R</i> : AFQT	<i>R</i> : Full Composite	ΔR	<i>R</i>	<i>R</i> : AFQT	<i>R</i> : Full Composite	ΔR
Career Intentions	.40	.08	.41	.33	.29	.27	.36	.09
Commitment	.40	.04	.42	.38	.46	.00	.48	.48

Note. Bolded values indicate statistical significance at $p < .05$; r = the bivariate correlation from the FIML correlation matrix.

Table 4.7. *Incremental Validity Results for the Full Performance Composite*

Performance Criteria	Analysis Sample				Holdout Sample			
	<i>r</i>	<i>R</i> : AFQT	<i>R</i> : Full Composite	ΔR	<i>r</i>	<i>R</i> : AFQT	<i>R</i> : Full Composite	ΔR
Leadership Potential	.28	.09	.32	.23	.24	.26	.34	.08
OCS APFT average	.56	.04	.56	.52	.51	.04	.52	.48
OCS History average	.04	.44	.45	.01	-			
OCS Total Score	.44	.18	.49	.31	.44	.25	.49	.24
OCS Leadership Score	.37	.09	.38	.28	.32	.22	.38	.16
OCS Fitness Score	.56	.00	.56	.56	.56	.03	.57	.54
OCS Academic Score	.04	.47	.48	.01	.11	.48	.49	.01

Note. Bolded values indicate statistical significance at $p < .05$; r = the bivariate correlation from the FIML correlation matrix.

Cross-Validation

The full composites were cross-validated using the holdout sample ($n = 246$). The full Continuance Composite provided significant incremental validity beyond AFQT for the prediction of Career Intentions ($\Delta R = .09$) and Affective Commitment ($\Delta R = .48$). The full Performance Composite provided incremental validity for the prediction of Peer Ratings of Leadership Potential ($\Delta R = .08$), OCS APFT ($\Delta R = .48$), OCS Total score ($\Delta R = .24$), OCS Leadership Score ($\Delta R = .16$), and OCS Fitness Score ($\Delta R = .54$). The patterns of prediction were similar for the analysis sample and the holdout sample. The most obvious difference in the two

samples was the prediction of AFQT with Career Intentions and with Peer Ratings of Leadership Potential. The validity coefficients were higher in the holdout sample, but both were non-significant. However, the higher validity coefficients led to lower ΔR values for both criteria. This discrepancy is likely due to slight variations between the two samples, including the smaller sample size in the holdout sample. Notably, the bivariate correlations between the full composites and criteria, independent of AFQT, are very similar for the analysis and holdout samples. The exception to this is in the prediction of Career Intentions (see Table 4.6).

Subgroup Differences

To examine subgroup differences of the composite scores, we first computed the means and standard deviations for each composite on the subgroups of interest (i.e., gender, race, and ethnicity). We then determined the extent to which the means differ using an independent samples *t*-test and a Cohen's *d* (Cohen, 1988) (see Chapter 2 for more details on this approach). The results of this analysis are presented in Table 4.8. There were no significant mean differences between the composites for Black and White officer candidates and for Hispanic and White, non-Hispanic officer candidates. There were significant mean differences between Male and Female officer candidates on the RBI Performance Composite ($d = -.61$) and the Total Performance Composite ($d = -.44$) such that Males scored higher than Females. This is not surprising given the strong weight of physical fitness scales – which tend to favor males – in the two performance composites (see Table B.3 in Appendix B and discussion in Chapter 2).

These results are limited by the small sizes of the minority samples. In particular, very few Hispanic ($n = 18$) and Black ($n = 53$) officer candidates were included in the analyses. Subgroup differences of the composites scores should be examined in a larger sample to determine the generalizability of these findings. Based on these preliminary findings the continuance composites did not exhibit any subgroup differences for minority groups. The performance composites show differences for females, but not Hispanics or Blacks.

Conclusion

We constructed two composites for predicting both performance and continuance criteria. The first included only RBI scales, whereas the second included the RBI scales and scales from TAPAS, LKT, Work Values, and Army Identity Structure measures found to provide incremental prediction of key criteria. Both composites yielded significant incremental validity beyond AFQT for EOC and administrative criteria while minimizing the number of component scales. The RBI composite offers significant predictive utility and relies on the administration of only one measure. The addition of the predictors included in the full composite yields only modest additional incremental validity to the prediction of the criteria. The cross-validation of the full composites in the holdout sample shows that the composites may be generalizable to independent samples and have the potential to be useful prediction tools for OCS. However, the performance composites also yielded large gender subgroup differences.

Table 4.8. *Subgroup Differences on Continuance and Performance Composites for the Enlistment Option Cohort*

Composites	Gender					Ethnicity					Race				
	Male (M)		Female (F)		M-F	White, Non-Hispanic (WNH)		Hispanic (H)		WNH-H	White (W)		Black (B)		W-B
	<i>n</i> = 269 - 616		<i>n</i> = 40 - 91			<i>n</i> = 225-508		<i>n</i> = 7 -18			<i>n</i> = 252-575		<i>n</i> = 25-53		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>
RBI Continuance	.12	.75	-.03	.84	-.20	.12	.77	.12	.83	.00	.11	.77	.28	.82	.22
RBI Performance	.21	.71	-.23	.74	-.61	.19	.74	.01	.52	-.28	.17	.74	.31	.71	.20
Total Continuance	-.01	.66	-.03	.62	-.03	-.01	.66	.06	.74	.10	-.01	.66	.08	.67	.14
Total Performance	.08	.66	-.20	.60	-.44	.10	.66	-.04	.60	-.22	.08	.66	-.01	.70	-.13

Note. $d = (M_{\text{COMPARISON}} - M_{\text{REFERENT}}) / SD_{\text{POOLED}}$, where the referent group is the majority group (Male, White, Non-Hispanic, and White). Results are based on the raw data combining the Phase 2 and Phase 3 samples. Bolded values indicate statistical significance at $p < .05$. Significance is based on the independent samples t -test difference between the two means.

Chapter 5: Conclusions and Next Steps

The purpose of this chapter is to summarize the previous four chapters, with particular emphasis on the results presented in Chapters 3 and 4. We begin with a brief summary of the objectives of this research and the methodology. We then summarize the results of the data analyses and the conclusions we can draw from them. This is followed by a discussion of the practical and theoretical implications of the current research as well as its limitations. Finally, we discuss future steps for continuing this program of work.

Background and Objectives

In Chapter 1, we learned that in previous years OCS had administered measures other than the ASVAB to predict officer performance and continuance. However, the current framework for selecting OCS candidates does not allow for “screening out” applicants with low potential for performance or continuance, nor for “screening in” high-potential individuals in a systematic way. Instead, the current system relies on subjective evaluations that may be susceptible to errors and biases. In response to this, ARI initiated a research program called “SelectOCS.” SelectOCS began with the development of multiple “non-cognitive” measures to predict officer performance and continuance, with particular emphasis on applicants entering OCS through the newer enlistment option (Russell & Tremble, 2011). Initial results were promising, with the OBEF predicting key outcomes three years after the initial administration (Allen & Young, 2012). In SelectOCS Phase 2, a new OBEF that measured aspects of personality, work values, motivation, self-efficacy, implicit leadership, commitment, and identity was piloted and found to predict key outcomes (Russell et al., 2011). These non-cognitive domains are thought to include both distal and proximal predictors of leader performance and officer continuance (see Figure 1 in Chapter 1).

The current “Phase 3” report continues the SelectOCS research program by accomplishing three objectives: (a) cross-validate the results from SelectOCS Phase 2 with a sample of newly accessioned OCS candidates, (b) select the most promising individual instrument for predicting officer performance and continuance, and (c) develop a set of empirical selection composites that balances both prediction and parsimony. To accomplish these objectives, we executed a longitudinal validation research project. We administered the OBEF to all candidates in five OCS classes at the end of their first week. We then collected criterion data through a second, end-of-class (EOC) survey and from OCS administrative records. We obtained candidates’ ASVAB scores from Army personnel records. The data were rigorously cleaned in preparation for analysis.

Summary of Results

The results summary is organized by the three objectives described above. We addressed the first and second objectives in Chapter 3, while we addressed the third objective in Chapter 4.

Objective 1: Cross-Validate Phase 2 Results

The demographic, OBEF, and performance variables differed substantially between the Phase 2 and Phase 3 samples. The Phase 3 sample had a much higher rate of enlistment option candidates (72% versus 56%) than Phase 2, more advanced degrees, and higher AFQT scores. Although the distribution of demographics across the Phase 2 and Phase 3 samples was relatively consistent, Phase 3 had slightly more white, non-Hispanic males and tended to be younger. This is consistent with the increased representation of enlistment option candidates, who tend to be young, white, and non-Hispanic. In terms of OBEF composition, where differences existed, the Phase 3 sample tended to have significantly higher scores than the Phase 2 sample on positively valenced scales (e.g., RBI Achievement) and lower scores on negatively valenced scales (e.g., RBI Equity Sensitivity). However, there were exceptions to this trend, such as with the TAPAS Curiosity and Intellectual Efficiency scales, where Phase 2 candidates scored higher than Phase 3 candidates. The Phase 3 sample also tended to have higher scores on the key performance and continuance criteria, such as Affective Commitment and Peer Ratings of Leadership Potential. This led us to question past reliance on scores standardized within class, as standardization will mask any differences between samples.

Despite the differences between the Phase 2 and Phase 3 samples, results of the multi-group measurement equivalence analyses suggest that, for the majority of the OBEF instruments, the predictive validity results were very consistent across the two samples. However, there was a significant difference in how the RBI predicted candidate physical fitness, as measured by their APFT scores. In general, the relation between the RBI and APFT was even stronger in Phase 3 than in Phase 2, suggesting that the results found in Phase 2 were not overestimates. In summary, these results suggest that the OBEF predicted key officer outcomes to a similar degree in SelectOCS Phase 2 and Phase 3, despite differences in the samples. These results provide support for using the OBEF operationally and allowed us to combine the Phase 2 and Phase 3 samples for subsequent analyses.

Objective 2: Select the Most Promising OBEF Instrument

The second objective was to select one instrument administered in the OBEF for use in future research based on its potential for predicting key performance and continuance outcomes of interest. Note that this analysis did not account for other factors that might be relevant to this task, such as administration time, adverse impact, and susceptibility to faking or coaching. Results of the incremental validity analyses suggest that the RBI is the most promising single measure for predicting key outcomes in both the in-service and enlistment option samples. The next most promising measure overall was the TAPAS, followed by the Work Values and Army Identity Structure scales. We therefore selected the RBI as the base instrument for developing a set of empirical selection composites.

Objective 3: Develop Empirical Composites

The final task in the present effort was to develop a set of empirical selection composites. Because there is limited time in an operational setting, we wanted the set of composites developed to be maximally predictive of key outcomes, yet to comprise as few scales as possible. In general, the OBEF scales that predicted continuance criteria (Affective Commitment, Career

Intentions) differed from the OBEF scales that predicted performance criteria (Academic, Leadership, Fitness). Therefore, we developed separate composites for each outcome category. Given the small in-service sample and their diminishing role in OCS, we only developed composites for candidates who entered OCS through the enlistment option. We created these composites by (a) developing initial RBI composites, (b) identifying additional scales to add to the RBI composites, (c) constructing final composites, and (d) evaluating the final composites.

The new composites accounted for an additional 7.8% to 31.4% of the variance beyond that afforded by the AFQT on all key outcomes of interest except Academic performance. Most importantly, the validity decreased little in a separate holdout sample, suggesting that these composites are stable predictors of key outcomes. The exception to this was the prediction of Career Intentions, where the ΔR decreased from .33 in the analysis sample to .09 in the holdout sample. However, this was primarily due to an increase in the predictive efficacy of the AFQT in the holdout sample.

In summary, these results suggest that the new OBEF composites would significantly enhance the current system for selecting OCS candidates through the enlistment option program.

Implications

For Practice

As described earlier, ARI is planning to administer some portions of the OBEF to applicants, in preparation for eventual operational use. The results presented here suggest that the OBEF holds considerable promise for identifying applicants with high and low potential for performance and for staying in the Army beyond their initial ADSO. To illustrate the potential operational utility of the OBEF, we constructed the empirical composites constructed in Chapter 4 on the Phase 3 enlistment option sample. We then segmented scores on the two composites into thirds (top, middle, and bottom). Finally, we compared the three groups on two key outcomes of interest: (a) candidates' current active duty career intentions and (b) their average OML rank. The results are presented in Figure 2.

The figures demonstrate that applying either of these composites operationally could have a positive effect on selection. With the Performance Composite, applicants in the top third are ranked 12 spots higher (on average) on the OCS OML than those in the middle third, and 25 spots higher (on average) than those in the bottom third. Meanwhile, nearly half of the candidates in the bottom third of the OBEF continuance composite indicated they were either undecided about staying beyond their initial ADSO or intended to leave after their initial ADSO. In contrast, only 13% of those in the top third of the OBEF composite intended to leave after their ADSO. These results suggest that either of these composites could be used to (a) identify high-potential OCS applicants (e.g., by selecting the top third) or (b) screen out OCS applicants with little performance or continuance potential. Alternatively, the Army could use the two composites as multiple hurdles for selecting high-potential OCS candidates. As OCS becomes more selective with the reduction in force requirements, these composites could provide substantial operational benefit to the Army.

However, the Army would need to make a number of other additional decisions to use these composites operationally. First, the method of weighting or combining the two composites will need to be determined. For example, the Army could use a compensatory system where the applicant needs to reach a total combined score. Another option would be to have separate cut scores for each composite (i.e., a non-compensatory model). The potential adverse impact (particularly for gender) of these new composites will also need to be considered more closely when it comes to operational use of the OBEF.

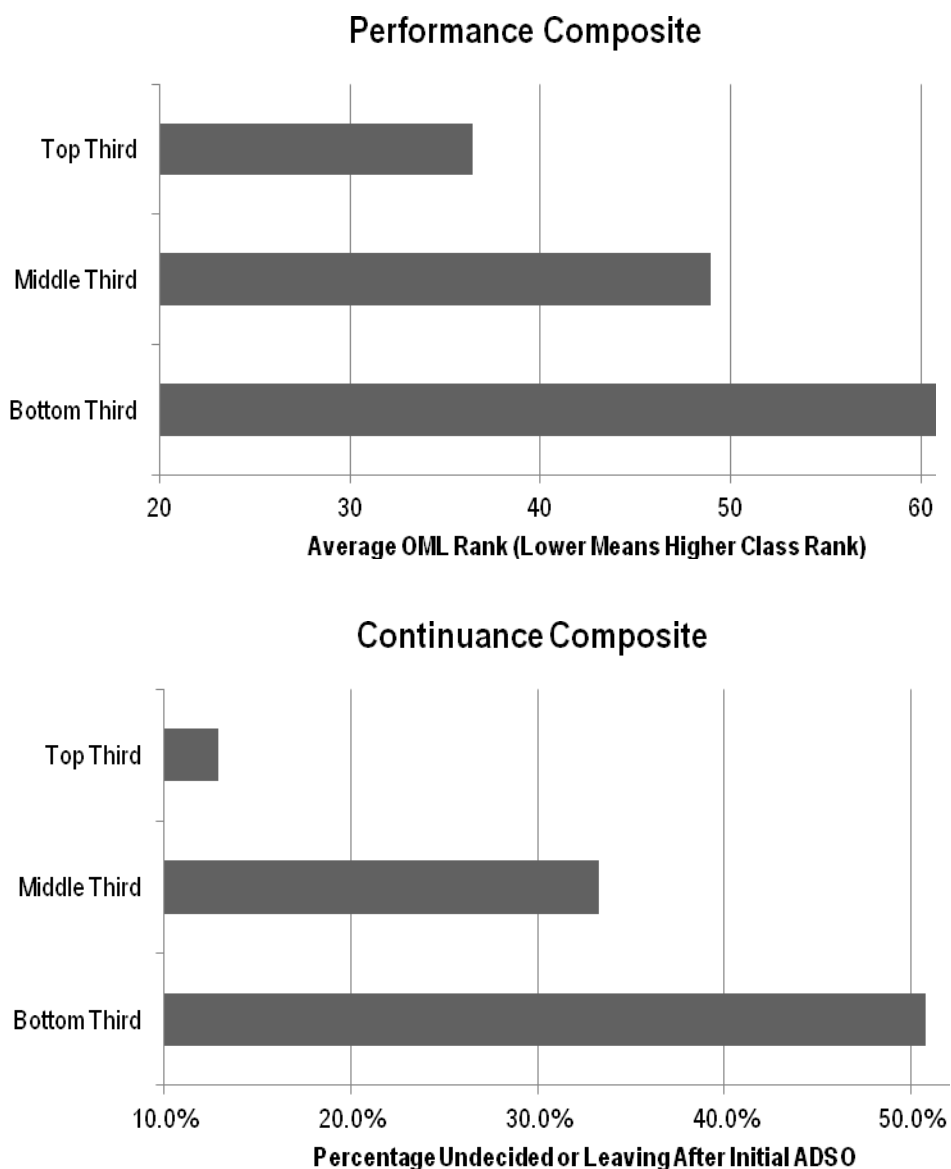


Figure 2. Expectancy charts of operational application of OBEF composites.

For Research

In Chapter 1, when describing how to predict key performance and continuance outcomes of interest, the predictor space was described in terms of cognitive ability, personality, work values, motivation, self-efficacy, implicit leadership, commitment, and identity structure. The OBEF assessed all of these aspects of the predictor space save cognitive ability. With the exception of implicit leadership as measured by the LKT (which did not relate to OCS Leadership Performance or Peer Ratings of Leadership Potential), all of the tested domains related to the theoretically expected outcomes of interest. However, the relative importance analyses suggest that some of these domains are more critical than others in a selection context. From the results presented in Chapters 3 and 4, we can conclude the following:

1. *Consistent with previous research, general cognitive aptitude is a strong predictor of technical aspects of job performance, but less predictive of interpersonal aspects of performance and continuance.* This is evidenced by the strong relation between AFQT and academic performance criteria in both the in-service and enlistment option samples. However, we should note that AFQT also predicted Peer Ratings of Leadership Potential in the enlistment option sample.
2. *For OCS candidates with no prior service (i.e., enlistment option), commitment and identity structure were the strongest predictors of continuance, followed by work values and motivations.* Examples of strong predictors of continuance include the RBI Affective Commitment scale, the Army Identity Structure scales, and the RBI Goal Expectations scale.
3. *For OCS candidates with no prior service (i.e., enlistment option), motivations and personality were the primary drivers of overall performance.* In particular, motivation for physical fitness and conditioning was a very strong predictor of performance in OCS. Although not nearly as strong relative to fitness motivation, aspects of personality such as TAPAS Achievement and RBI Emotional Stability also predicted performance outcomes, particularly Peer Ratings of Leadership Potential.
4. *Predictive validity results for OCS candidates with prior service (i.e., in-service option) were similar to the results found for non-prior-service candidates (i.e., enlistment option) with a few key exceptions.* For a number of reasons, we did not examine the relationship between the OBEF and key criteria as rigorously in the in-service sample as we did for the enlistment option sample. However, the findings reported in Chapter 3 suggest that the results are consistent for the two samples. For example, aspects of commitment and Army Identity Structure were strong predictors of Affective Commitment and Career Intentions, whereas fitness motivations were highly predictive of performance. However, there were a few key differences. Most notably, RBI Hostility to Authority, RBI ARC Hostility to Authority, and RBI Equity Sensitivity, which can be thought of as aspects of an “emotional stability” or “adjustment” dimension of personality, all strongly related to continuance outcomes. These scales were unrelated to the same outcomes in the enlistment option sample. The relation between some work values scales and key outcomes was also different in this sample. For example, the Work Values scales Benevolence and Structure related to Peer Ratings of Leadership Potential in the in-service sample, but not in the enlistment option sample.

In summary, these results suggest that this effort has theoretical as well as practical implications. In some cases, the results support previous research, but they also expand on it by examining the relative importance of various predictors.

Limitations

By collecting data early in their tenure at OCS and using a longitudinal design, we attempted to make this research as close to an applicant context as possible. Despite these attempts, there were a number of differences between the respondents, the setting, and the administration instructions in this research when compared to an operational setting. For example, the respondents in this sample have already gone through screening to gain admission to OCS, which means the sample population in this research will be more restricted on key predictor constructs (e.g., cognitive ability, personality) than an applicant population. In terms of administration instructions, respondents in this research were told to answer as honestly as possible and that there were “no right or wrong” answers. Thus, respondents in this sample were less motivated to respond in a socially desirable manner, which could limit the generalizability of these results to an applicant population. Future research should account for these limitations.

Future Directions

As of this writing, plans are currently underway for ARI to test the OBEF in an applicant setting. In this project, ARI would administer a shortened version of the OBEF to OCS enlistment option applicants, most likely at a Military Entrance Processing Station (MEPS). A research project such as this would go a long way towards addressing the generalizability concerns expressed above and thus is a logical extension of the current research. It would provide a good setting to work out the operational issues described previously, such as the method of weighting the two composites and tradeoff between validity and adverse impact.

Another possibility is to conduct a separate faking tryout, where researchers ask participants to respond to the OBEF as if they were applicants, to see if the pattern of results changes between “applicant” instructions and “answer honestly” instructions. This avenue may be useful if the aforementioned project is delayed. In addition to these new projects, future research should also follow participants in Phase 2 and Phase 3 samples to see if the OBEF predicts in-unit performance and continuance, not just training performance and continuance intentions. Although previous SelectOCS projects have suggested that the OBEF is a robust predictor of long-term outcomes (Allen & Young, 2012), the Phase 1 version of the OBEF used in that research project was quite different from the version administered to the samples under investigation here. Finally, future research should also more carefully consider other factors, such as adverse impact, in evaluating new predictors and using them operationally.

In addition to the research concerning the operational use of the OBEF described above, future research might also more closely examine how this research contributes to models of turnover and leader performance (e.g., Hom, 2011; Van Iddekinge et al., 2009). Researchers can accomplish this by modeling mediator and moderator relations, and by including certain covariates (e.g., prior service) in a more systematic way. Additionally, this research could be

used to determine whether the models found in previous turnover and leadership research apply to the prediction of key outcomes in an officer sample. In summary, future research can examine both the operational application of the OBEF and the implications this work has for contributing to our understanding of officer performance and continuance.

References

- Allen, M.T., Thibodeaux, C., & Babin, N.E. (2011). *Examining U.S. Army officer attitudes by commissioning source* (FR-10-66b). Alexandria, VA: Human Resources Research Organization.
- Allen, M.T., & Young, M.C. (Eds.) (2012). *Longitudinal validation of non-cognitive officer selection measures for the U.S. Army Officer Candidate School (OCS)* (Technical Report 1323). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Bureau of Labor Statistics (2005). *The NLS handbook*. Retrieved April 22, 2009, from <http://www.bls.gov/nls/handbook/nlshndbk.htm>
- Burket, G.R. (1964). A study of reduced rank models for multiple prediction. *Psychometrika Monograph Supplement*, No. 12.
- Campbell, J.P. (2012). Behavior, performance, and effectiveness—in the 21st century. In S. Kozlowski (Ed.), *The Oxford Handbook of Organizational Psychology*. New York: Oxford University Press.
- Campbell, J.P., & Kuncel, N.R. (2002). Individual and team training. In N. Anderson, D.S. Ones, H.K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology, Volume I: Personnel psychology* (pp. 278-312). Thousand Oaks, CA: Sage Publications, Ltd.
- Campbell, J.P., & Knapp, D.J. (Eds.) (2001). *Exploring the limits in personnel selection and classification*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, C.E. (1993). A theory of performance. In N. Schmitt & W. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass.
- Chan, K., & Drasgow, F. (2001). Toward a theory of individual differences and leadership: Understanding the motivation to lead. *Journal of Applied Psychology*, 86(3), 481-498.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, 2nd edition*. Hillsdale, NJ: Erlbaum.
- Colquitt, J.A., LePine, J.A., & Noe, R.A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85(5), 678-707.
- Connelly, M.S., Gilbert, J.A., Zaccaro, S.J., Threlfall, K.V., Marks, M.A., & Mumford, M.D. (2000). Exploring the relationship of leadership skills and knowledge to leader performance. *The Leadership Quarterly*, 11(1), 65-86.

- Costa, P.T., & McCrae, R.R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13(6), 653-665.
- Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, 105(2), 317-327.
- Department of Defense (DOD) Statistical Information Analysis Division (SIAD) (November 30, 2011). *Department of Defense active duty military personnel by rank/grade*. Retrieved January 26, 2012, from <http://siadapp.dmdc.osd.mil/personnel/MILITARY/rg1111.pdf>
- Drasgow, F., Stark, S., & Chernyshenko, O.S. (2006, November). *Toward the next generation of personality assessment systems to support personnel selection and classification decisions*. Paper presented at the 48th annual conference of the International Military Testing Association, Kingston, Ontario, Canada.
- Enders, C.K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, 61(5), 713-740.
- Enders, C.K. (2006). Analyzing structural equation models with missing data. In G.R. Hancock & R.O. Mueller (Eds.), *A second course in structural equation modeling* (pp. 313-344). Information Age: Greenwich, CT.
- Enders, C.K., & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8, 430-457.
- Epitropaki, O., & Martin, R. (2005). From ideal to real: A longitudinal study of the role of implicit leadership theories on leader-member exchanges and employee outcomes. *Journal of Applied Psychology*, 90(4), 659-676.
- Hendricks, J.W., & Payne, S.C. (2007). Beyond the big five: Leader goal orientation as a predictor of leadership effectiveness. *Human Performance*, 20(4), 317-343.
- Henning, C.A. (2006). *Army officer shortages: Background and issues for Congress*. Washington, DC: Congressional Research Service, The Library of Congress. <http://www.fas.org/sgp/crs/natsec/RL33518.pdf>
- Hom, P.W. (2011). Organization exit. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization* (pp. 325-375). Washington, DC: American Psychological Association.
- Hunter, J.E., & Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

- Johnson, J.W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35(1), 1-19.
- Johnson, J.W., & Lebreton, J.M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7(3), 238-257.
- Judge, T.A., Bono, J.E., Ilies, R., & Gerhardt, M.W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Experimental Psychology*, 87(4), 765-780.
- Judge, T.A., Heller, D., & Mount, M.K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, 87(3), 530-541.
- Karrasch, A.I. (2003). Antecedents and consequences of organizational commitment. *Military Psychology*, 15(3), 225-236.
- Kilcullen, R.N., White, L.A., Mumford, M.D., & Mack, H. (1995). Assessing the construct validity of rational biodata scales. *Military Psychology*, 7(1), 17-28.
- Knapp, D.J., & Heffner, T.S. (Eds.) (2009). *Validating future force performance measures (Army Class): End of training longitudinal validation* (Technical Report 1257). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D.J., & Heffner, T.S. (Eds.) (2010). *Expanded enlistment eligibility metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D.J., & Heffner, T.S. (Eds.) (2011). *Tier One Performance Screen initial operational test and evaluation: 2010 annual report* (Technical Report 1296). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Knapp, D.J., Sager, C.E., & Tremble, T.R. (Eds.) (2005). *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Kraut, A.I. (1975). Prediction of managerial success by peer and training-staff ratings. *Journal of Applied Psychology*, 60 (1), 14-19.
- Kraut, A.I. (2005, October). *A powerful and simple way to predict executive success: Results from a 25-year study of peer evaluations*. Paper presented at the Society for Industrial and Organizational Psychology's Leading Edge Consortium, St. Louis, MO.
- Lawley, D. (1943). A note on Karl Pearson's selection formulae. *Royal Society of Edinburgh, Proceedings, Section A*, 62, 28-30.

- Mallows, C.L. (1973). Some comments on Cp. *Technometrics*, 15(4), 661-675.
- McCloy, R., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999). *Development of the O*NET Paper and Pencil Work Importance Locator*. Raleigh, NC: National Center for O*NET Development.
- McDaniel, M.A., Psotka, J., & Legree, P.J. (2009, April). *Toward an understanding of situational judgment item validity*. Paper presented at the 24th annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- McHenry, J.J., Hough, L.M., Toquam, J.L., & Hanson, M.A. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43(2), 335-354.
- Meyer, J.P., Becker, T.E., & Van Dick, R. (2006). Social identities and commitments at work: Toward an integrative model. *Journal of Organizational Behavior*, 27(5), 665-683.
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 42, 431-462.
- Muthén, B.O., & Muthén, L.K. (2004). *Mplus user's guide*. Los Angeles: Muthén & Muthén.
- Oliver, J., Ardison, S., Russell, T.L., & Babin, N.E. (2011). *Identification and accessioning of individuals for the Officer Candidate School (OCS)* (Study Report 2011-01). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Oppler, S.H., McCloy, R.A., & Campbell, J.P. (2001). The prediction of supervisory and leadership performance. In J.P. Campbell & D.J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 389-409). Mahwah, NJ: Lawrence Erlbaum Associates Publishers, Inc.
- Paullin, C., Sinclair, S.L., Moriarty, K.O., Vasilopoulos, N.L., Campbell, R.C., & Russell, T.L. (2011). *Army officer job analysis: Identifying performance requirements to inform officer selection and assignment* (Technical Report 1295). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Paullin, C., Moriarty, K.O., Legree, P.L., Russell, T.L., Bynum, B., Ford, L., Sinclair, A., & Campbell, R.C. (2012). *Predictors and criteria for officer selection and classification* (in preparation). Alexandria, VA: Human Resources Research Organization.
- Pearlman, K., Allen, M.T., Putka, D.J., Hooper, A.C., & Waters, S.D. (2009). Situational judgment test. In T. L. Russell & T.R. Tremble (Eds.), *Development and validation of measures for selecting soldiers for the Officer Candidate School* (FR-09-38). Alexandria, VA: Human Resources Research Organization.

- Putka, D.J. (Ed.) (2009). *Initial development and validation of assessments for predicting disenrollment of four-year scholarship recipients from the Reserved Officer Training Corps (ROTC)* (Study Report 2009-06). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Putka, D.J., Kilcullen, R.N., Tremble, T., Wasko, L.E., & Shaw, M.N. (2009). *An investigation into the validity of the Cadet Background Experiences Forms* (FR 09-53). Alexandria, VA: Human Resources Research Organization.
- Putka, D.J., Le, H., McCloy, R.A., & Diaz, T. (2008). Ill-structured measurement designs in organizational research: Implications for estimating interrater reliability. *Journal of Applied Psychology*, 93(5), 959-981.
- Ree, M.J., Carretta, T.R., & Teachout, M.S. (1995). Role of ability and prior knowledge in complex training performance. *Journal of Applied Psychology*, 80(6), 721-730.
- Robbins, J.M., Allen, M.T., & Putka, D.J. (2011). Identity, affectivity, and values. In T.L. Russell & T.R. Tremble (Eds.), *Development and validation of measures for selecting soldiers for the officer candidate school* (Study Note 2011-02, pp. 25-39). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Roth, P.L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47(3), 537-559.
- Rumsey, M.G. (2012). Selecting Army officers. In P. Ramsberger, N. Wooten, & M.G. Rumsey (Eds.) *A history of the research into methods for selecting and classifying U.S. Army personnel, 1917-2009*. Lewiston, NY: Mellon Press.
- Russell, T.L., Allen, M.T., & Babin, N.E. (2011). *Measures for selecting Soldiers for the Officer Candidate School (OCS): Phase II* (FR-10-66a). Alexandria, VA: Human Resources Research Organization.
- Russell, T.L., & Tremble, T.R. (Eds.) (2011). *Development and validation of measures for selecting soldiers for the Officer Candidate School* (Study Note 2011-02). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Schafer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177.
- Shamir, B., & Kark, R. (2004). A single-item graphic scale for the measurement of organizational identification. *Journal of Occupational and Organizational Psychology*, 77(1), 115-123.

- Schleicher, D.J., Hansen, S.D., & Fox, K.E. (2011). Job attitudes and work values. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol 3: Maintaining, expanding, and contracting the organization* (pp. 137-189). Washington, DC: American Psychological Association.
- Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.
- Schmitt, N., & Ployhart, R. E. (1999). Estimates of cross-validity for stepwise regression and with predictor selection. *Journal of Applied Psychology*, 84(1), 50-57.
- Stark, S., Drasgow, F., & Chernyshenko, O.S. (October, 2008). *Update on Tailored Adaptive Personality Assessment System (TAPAS): The next generation of personality assessment systems to support personnel selection and classification decisions*. Paper presented at the 50th annual conference of the International Military Testing Association. Amsterdam, Netherlands.
- Strickland, W.J. (Ed.) (2005). *A longitudinal examination of first term attrition and reenlistment among FY 1999 enlisted accessions* (Technical Report 1172). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks-Cole.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology on intergroup relations* (pp. 7-24). Chicago: Nelson-Hall.
- Tett, R.P., & Meyer, J.P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on meta-analytic findings. *Personnel Psychology*, 46(2), 259-293.
- Tice, J. (2010, September 29). Army to dump captains after multiple passovers. *Army Times*. Retrieved Feb 22, 2012 from <http://www.armytimes.com/news/2010/09/army-up-and-out-for-captains-092710w/>
- Van Iddekinge, C.H., Ferris, G.R., & Heffner, T.S. (2009). Test of a multistage model of distal and proximal antecedents of leader performance. *Personnel Psychology*, 62(3), 463-495.
- Wardyniski, C., Lyle, D.S., & Colarusso, M.J. (2009). *Towards a U.S. Army officer corps strategy for success: A proposed human capital model focused upon talent*. Carlisle, PA: Strategic Studies Institute, U.S. Army War College. <http://www.strategicstudiesinstitute.army.mil/pubs/display.cfm?pubID=912>

- Wasko, L.E., Putka, D.J., Legree, P., & Kilcullen, R.N. (2011). *Validation of measures for predicting leader development and assessment course (LDAC) performance* (FR-11-19). Alexandria, VA: Human Resources Research Organization.
- White, L.A., Young, M.C., Hunter, A.E., & Rumsey, M.G. (2008). Lessons learned in transitioning personality measures from research to operational settings. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1(3), 291-295.

Appendix A

Manual for Collecting Data in CY2011 for SelectOCS Phase 3

1. Overview of Data Collections

The Army Research Institute (ARI) and the Human Resources Research Organization (HumRRO) are conducting projects designed to improve recruitment and selection for the Army's Officer Candidate School. The purpose of this manual is to standardize and facilitate data collection for the Select OCS 3 project. The data collections will involve a group administration of the Officer Background and Experience Form (OBEF) to cross-validate non-cognitive officer selection measures. The first section describes the standard operating procedures for administering the OBEF and the final section provides information on reporting.

2. OBEF Administration

Packing List

For each class, we will need a total of 150 packets, 75 of set A and 75 of set B. Set A packets will always have an odd numbered identification number, and Set B will always have an even identification number. For each, you will need:

75 manila envelopes, with an ID code and set letter (A) on the outer label. Each packet for set A should contain the following:

1. Privacy Act Statement (1 pg)
2. Informed Consent Form (2 pgs single-sided, stapled, and pre-slugged with the identification numbers)
3. Demographic Form (pre-slugged with identification numbers)
4. OBEF Form 1-A (printed double sided and stapled)
 - a. Scantron should be paper clipped to Form 1-A (pre-slugged with identification number and the Form letter/number)
5. OBEF Form 2-A (printed double sided and stapled)
 - a. Scantron should be paper clipped to Form 2-A (pre-slugged with identification number and the Form letter/number)

75 manila envelopes, with an ID code and set letter (B) on the outer label. Each packet for set B should contain the following:

1. Privacy Act Statement (1 pg)
2. Informed Consent Form (2 pgs single-sided, stapled, and pre-slugged with the identification numbers)
3. Demographic Form (pre-slugged with identification numbers)
4. OBEF Form 1-B (printed double sided and stapled)
 - a. Scantron should be paper clipped to Form 1-B (pre-slugged with identification number and the Form letter/number)
5. OBEF Form 2-B (printed double sided and stapled)
 - a. Scantron should be paper clipped to Form 2-B (pre-slugged with identification number and the Form letter/number)

You will also need supplies, including:

- 400 sharpened #2 pencils
- Pencil Sharpener
- Test session records/problem log sheets (see Attachment A)
- 10 Clipboards (optional, for examinees who do not have desks when the room is too full)
- Laptop

Overview of OBEF Administration Steps

~ Allow yourself 1 to 1½ hours to set up the room before the session.

- A. If the chairs are on top of the tables, turn them over and push them all the way under the table. (This will allow the Candidates to squeeze through the long rows easily.)
- B. Place a manila envelope and two #2 pencils on the table in front of each chair. Be sure the manila envelopes are in order according to the candidate ID number.
- C. Place ~10 sets of manila envelopes, #2 pencils (two per set), and clipboards in the front of the room. Be sure the envelopes are in order according to the candidate ID number. These items will be for any individuals who have to sit on the floor if seating is insufficient.
- D. Place extra materials in the front of the room where you can access them easily, just in case if one of the packets is missing any materials.
- E. When the applicants arrive (likely in large groups) be sure they fill up the spaces in blocks. (Other blocks of seats can be reserved for late comers.)
- F. Read the introduction script starting on page 5 aloud to examinees. It involves giving instructions for the:
 - a. Privacy Act and Informed Consent Forms
 - b. Demographic Form
 - c. OBEF booklet and OBEF answer sheets
- G. If everyone arrives at once and one Test Administrator (TA) is free, the TA who is not speaking can pick up Consent Forms. Check that:
 - a. The AKO email address and the date of birth information is legible, the consent form is signed, and the two boxes at the top (i.e., over 18 and voluntary participation) are checked.
 - b. After checking, place them in numerical order by Form and Candidate ID code.
 - c. Enter information from the informed consent forms into a spreadsheet using the class rosters as a starting point.
- H. For individuals arriving late, ask them to wait outside until you have a large group of individuals (~10) and then have the other administrator go through the script so they can “catch up” with the rest of the group.

- I. Circulate through the room and make sure examinees are bubbling in items appropriately. The space in the room is tight and you will not be able to walk down the rows.
- J. During the session, make an announcement reminding participants that they are to complete both of the forms in their packets as well as the demographic form.
- K. As Candidates start to leave, ask them to hand their completed packets to you and to please put their chairs on top of the tables if they are the last group that will be using the room that day.
- L. Place boxes in the front of the room and the empty pencil case holder. One box should be for collecting unused test forms, the second box for completed Form As and the third box for completed Form Bs.
- M. As Candidates turn in their completed materials, check to be sure that they have completed all the forms.
- N. After the testing session:

Objectives for preparing the test materials

- 1. Batch the consent forms and make sure they are separated in two piles by Form and ordered according to Candidate ID.
- 2. Batch the reusable OBEF Forms and arrange according to form and erase any stray marks.
- 3. Order the manila envelopes in the order of the Candidate ID (Be sure to keep Form A and B separate).
- 4. To complete the processing of the forms:
 - Open Form A manila envelopes. Create four piles. All piles should be ordered according to Candidate ID code:
 - Form A consent forms
 - Demographic form
 - Scantron 1-A
 - Scantron 2-A
 - Open Form B manila envelopes. Create four piles. All piles should be ordered according to Candidate ID code:
 - Form B consent forms
 - Demographic Form
 - Scantron 1-B
 - Scantron 2-B.
 - Check that the Candidate used the correct OBEF form (A or B) with the corresponding pre-slugged scantron by checking that:
 - 1A and 2 B have 156 questions (the first 96 questions have a T/F response option).
 - 1B and 2A have 144 questions (the first 132 questions have a 5-pt response option).
 - Check for stray marks and/or other potential problems on the scantron.
 - Attach your completed Test Session Record (from Attachment A) to the batched scantrons and demographic forms.

- Throw away additional Privacy Statements, the top page of consent forms, and/or empty manila envelopes.
- O. Materials with personally identifying information (Name, Date of Birth) cannot be checked in your bag or shipped. They must be hand-carried to HumRRO's offices and given to project staff. If completed correctly, this should only include the Informed Consent forms.

OBEF Session Script

[Be sure to talk loudly so the people in the back of the room can hear. Do not read this script verbatim, but be sure to describe all of the key points]

1. Welcome the participants and introduce yourselves

Good morning (afternoon)! Thank you for joining us today. I am _____ and this is _____. We represent the U.S. Army Research Institute for the Behavioral and Social Sciences. ARI is essentially the Army's research lab. ARI is the resource that the Army turns to in order to investigate topics regarding leadership, training, selection, and retention.

2. Describe the purpose of the project

The purpose of this project is to develop and evaluate measures that could be used to improve selection for OCS.

3. Provide an overview of the task

Today you've been brought here to complete two questionnaires. Your task is very simple but extremely important in our research. The quality of our recommendations to the Army are dependent on your participation today. If you are careless, fail to listen to instructions, or give very little thought when reading the questions, that provides very little or lower quality information for the Army. So it is important that you respond as honestly as possible to these questions and answer all the questions asked in both questionnaires.

4. Walk through forms in the packet

Placed in front you is a packet with the materials. Open the packet but make sure you keep the materials in the order as it is in the envelope.

*On the top page is the **privacy act**. (Hold up the form). This tells you the purpose of the study and information regarding the confidentiality of the information you provide today. **That is, your responses today WILL NOT go into your record and will be used for research purposes ONLY.** So again we emphasize that you please take the time to respond as honestly as possible.*

*The next page is the **Informed consent**. (Hold up the informed consent). This document is similar to the Privacy Act, however it provides more detail regarding your participation such as the specific risks and benefits. Go ahead and tear off the top of this form (Tear off the top of the form) but keep the bottom of the informed consent on the table. We will go through that in a moment. The top of the informed consent and the Privacy Act are yours to keep, so go ahead and place those materials in your pocket or bag now.*

*Next we ask that you fill out the second page of the **Informed Consent**. (Hold up pg 2 of the informed consent). Does everyone have this page? Your signature on this page verifies that you are older than 18 years old and that you voluntarily are participating in this study. Please neatly write down your e-mail address and date of birth. Also, sign and date the informed consent form. It is important we are able to read this information, as we will need to contact you in the future.*

If asked about whether this survey is really voluntary. Indicate that it is and say *If you choose not to participate, please sit quietly in your seat until we dismiss you, so that you do not disturb the other Candidates.*

We are asking for your date of birth because this is a longitudinal study. We are collecting this information to link your current survey responses the information to two surveys that you will be taking in the future when you finish OCS and BOLC-B. We ultimately want to see how well your scores today predict your future performance.

Each of you has been assigned an identification code. Once we take this information from this location back to our offices, we swap out your date of birth with the ID codes. We take precautionary measures to ensure your privacy is maintained, however if you have any questions on this matter please refer to the information provided on the Informed Consent form to contact one of the POCs from ARI.

When you finish your Informed Consent form, please hand it towards the middle aisle and we will come by and pick it up. [To the extent possible, administrators should circulate through the room and look for problems in form completion prior to collection.]

*Now pull out the **demographic** form (Hold up the demographic form). Go ahead and mark down your information. After you are finished filling out the Demographic information go ahead and place it back into the manila envelope.*

*Now you should only have **two booklets with attached scantron forms** (Hold up the two booklets and scantron forms). Make sure you use the scantron that is attached to the booklet. You must fill out both of these forms. **PLEASE NOTE THAT WE ASK THAT YOU DO NOT FILL IN YOUR SOCIAL SECURITY NUMBER ON THE SCANTRON FORMS.** You do not need to bubble in anything except your responses to the questions.*

*At the end of the session when you are finished taking both tests, place all of your materials in the manila folders. Also please **DO NOT** paper clip any of your materials. You can just leave the paperclips on the table. Then let one of us know so that we can check your packet.*

5. **Tell them what to do when they are finished (this can vary somewhat by class, ask the company commander or cadre member that brings the group to the session for guidance).**

Please be respectful of your peers and work quietly at your desk. Are there any questions? Okay you may go ahead and begin.

Completing the Test Session Record

Complete a Test Session Record (Attachment A) each time you administer the OBEF. It is very important that you record the date and range of ID codes used in your session.

When recording a question or event, record the examinee's ID code and describe the question or incident. If the examinee had a question about a particular test item, record the test item number. Record any other unusual events (e.g., examinee falling asleep, fire alarm sounding). You should be sure to record any information that might be helpful in later examining the data. For example, if someone is completing the test very quickly and you suspect that he might be responding randomly, you should record that information on the log.

3. Reporting

Your Trip Report

The lead data collector on each trip is responsible for preparing a trip report. A template for the trip report appears in Attachment B.

Attachment A: Test Session Record

For Beginning of Class (BOC) Data Collections

SITE : _____

TEST DATE: _____ ID CODE RANGE: _____

TOTAL NUMBER OF PARTICIPANTS: _____

START TIME: _____ END TIME: _____

RECORD PREPARED BY: _____

PAGE _____ of _____ *(note: you can't fill in the second space until end of session)*

Include ID numbers of the individuals involved. Be sure to describe what you did in response (if anything) to the event. Please write legibly!

Attachment B. Trip Report Template

Location:

Date of Visit:

Data Collection Team:

Army POC:

Report prepared by:

Schedule:

Summary of the Sessions:

Provide a general description of the sessions.

Attach test session logs and any other additional notes.

Appendix B

Descriptive Statistics and Subgroup Differences

Table B.1. Predictor Descriptive Statistics for the SelectOCS Phase 3 Sample

Scale/Predictor	Total					α
	<i>N</i>	Min	Max	<i>M</i>	<i>SD</i>	
AFQT	452	32.00	99.00	86.70	10.76	--
Rational Biodata Inventory (RBI)						
Peer Leadership	446	2.17	5.00	3.98	0.54	.69
Achievement	446	1.75	5.00	4.20	0.48	.54
Fitness Motivation	446	1.88	5.00	3.94	0.58	.74
Hostility to Authority	446	1.00	4.33	1.87	0.49	.58
Generalized Self-Efficacy	446	2.50	5.00	4.33	0.48	.76
Affective Commitment	446	2.11	5.00	4.02	0.58	.75
Stress Tolerance	446	1.00	5.00	3.29	0.52	.68
ARC Hostility to Authority	446	1.00	5.00	2.82	0.64	.64
Interest in Leadership	446	1.25	5.00	3.97	0.70	.76
Leader Self-Efficacy	446	2.50	5.00	4.12	0.45	.78
Equity Sensitivity	446	1.00	4.33	2.29	0.57	.66
Tolerance for Injury	446	1.00	5.00	3.77	0.61	.64
Emotional Stability	446	1.60	5.00	3.64	0.69	.71
Goal Expectations	446	2.00	5.00	3.59	0.60	.67
Narcissism	446	1.00	5.00	2.79	0.79	.77
Tolerance for Ambiguity	446	1.50	4.25	2.79	0.48	.68
Social Acuity	446	1.63	5.00	3.78	0.55	.76
Work Values						
Benevolence (Likert)	446	2.20	5.00	4.24	0.54	.73
Social Work Environment (Likert)	446	1.80	5.00	3.82	0.61	.69
Selfless Service (Rank)	413	-2.28	3.46	0.83	1.16	--
Leadership (Rank)	413	-2.88	3.46	0.19	1.20	--
Recognition (Rank)	413	-2.88	3.46	0.00	1.08	--
Pay (Rank)	413	-3.46	3.46	0.08	1.04	--
Structure (Rank)	413	-2.88	2.88	-0.10	1.04	--
Comfort (Rank)	413	-2.88	2.88	-0.24	1.07	--
Home (Rank)	413	-2.54	2.88	-0.13	0.99	--
Challenge (Rank)	413	-2.88	2.54	0.01	0.96	--
Self-Direction (Rank)	413	-3.46	2.28	-0.11	0.90	--
Teamwork (Rank)	413	-3.46	3.46	-0.12	0.87	--
Variety (Rank)	413	-2.88	2.54	0.11	0.87	--
Tailored Adaptive Personality Assessment System (TAPAS)						
Achievement	434	-1.91	1.93	0.32	0.62	--
Curiosity/Continuous Learning	434	-2.43	1.97	-0.03	0.69	--
Non-delinquency	434	-2.03	1.70	-0.15	0.58	--
Dominance/Leadership	434	-1.93	1.67	0.09	0.60	--
Even Temper	434	-1.86	1.96	0.14	0.57	--
Intellectual Efficiency	434	-1.87	1.94	0.11	0.72	--
Adjustment	434	-1.67	1.87	-0.07	0.67	--
Physical Conditioning	434	-1.39	2.60	0.52	0.77	--
Responsibility	434	-1.58	1.95	-0.10	0.61	--
Tolerance	434	-2.51	1.83	-0.57	0.85	--
Trust/Cooperation	434	-1.95	1.08	-0.58	0.53	--
Optimism	434	-1.69	2.02	0.20	0.68	--
Leader Knowledge Test (LKT)						
Traits	429	8.76	9.62	9.43	0.13	.78
Skills	424	8.78	9.62	9.40	0.13	.56
Army Identity Structure						
Overlap	447	1.00	7.00	5.45	1.13	--
Concept	447	1.00	7.00	4.33	1.48	--
Conflict	446	1.00	7.00	5.06	1.37	--

Table B.2. Descriptive Statistics and Intercorrelations for the SelectOCS Phase 3 End-of-Class and OCS Criterion Measures

	<i>N</i>	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11
1. Career Intentions	294	3.10	0.89	--										
2. Affective Commitment	294	3.94	0.72	.31	(.87)									
3. Continuance Commitment	294	3.01	0.96	.01	.17	(.82)								
4. Morale	293	4.18	0.72	.21	.36	-.15	--							
5. Branch Satisfaction	289	4.43	0.99	.09	.22	.03	.06	--						
6. Peer Ratings of Leadership Potential	410	5.74	1.40	.11	.05	-.06	.06	.03	(.75 ^a)					
7. OCS Total Order of Merit List Score ^b	315	-0.01	1.00	.11	.06	-.18	.01	.15	.46	--				
8. OCS Leadership Score ^a	315	-0.01	1.00	.16	.08	-.17	.00	.07	.48	.94				
9. OCS Academic Performance Score ^a	315	-0.00	1.00	-.07	.05	-.07	-.04	.29	.22	.54	.36			
10. OCS Average History Course Score	124	85.09	9.59	-.12	.19	.16	-.03	.35	.14	.34	.16	.76		
11. OCS Physical Fitness Score ^a	315	-0.00	0.99	.02	-.04	-.15	.02	.14	.19	.67	.46	.22	-.01	
12. OCS Average APFT Score	124	279.95	16.02	.16	-.08	-.24	-.14	.19	.31	.61	.42	.25	.13	.88

Note. Coefficients in bold are statistically significant $p < .05$. Values along the diagonals are coefficient alpha reliability estimates.

^aReliability estimate is the average interrater reliability estimate for the company-grade and field-grade ratings.

^bScore is standardized within class.

Table B.3. OBEF Subgroup Differences for the SelectOCS Phase 3 Sample

Scale/Predictor	OCS Option					Gender				
	In-Service (I)		Enlistment (E)		I-E	Male (M)		Female (F)		M-F
	<i>n</i> = 31 – 36		<i>n</i> = 302 – 326			<i>n</i> = 344 – 376		<i>n</i> = 46 – 54		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>
AFQT	75.72	17.51	88.79	8.39	0.75	87.16	10.26	85.11	11.70	-0.20
Rational Biodata Inventory (RBI)										
Peer Leadership	3.84	0.62	4.02	0.53	0.28	3.95	0.54	4.20	0.54	0.47
Achievement	4.09	0.65	4.23	0.45	0.23	4.18	0.48	4.39	0.46	0.44
Fitness Motivation	3.87	0.66	3.94	0.58	0.11	3.99	0.56	3.63	0.56	-0.65
Hostility to Authority	1.79	0.56	1.89	0.49	0.18	1.88	0.50	1.71	0.38	-0.35
Generalized Self-Efficacy	4.35	0.44	4.34	0.47	-0.01	4.31	0.49	4.45	0.41	0.29
Affective Commitment	4.01	0.55	4.02	0.59	0.01	4.02	0.58	4.02	0.65	0.00
Stress Tolerance	3.28	0.52	3.28	0.53	0.01	3.30	0.52	3.17	0.47	-0.25
ARC Hostility to Authority	2.64	0.67	2.84	0.63	0.30	2.84	0.65	2.75	0.60	-0.14
Interest in Leadership	3.95	0.71	3.95	0.71	0.01	3.97	0.68	3.99	0.85	0.03
Leader Self-Efficacy	4.18	0.43	4.11	0.45	-0.17	4.12	0.45	4.11	0.48	-0.02
Equity Sensitivity	2.30	0.58	2.30	0.57	-0.01	2.30	0.56	2.17	0.59	-0.24
Tolerance for Injury	3.54	0.65	3.82	0.61	0.43	3.80	0.57	3.54	0.77	-0.46
Emotional Stability	3.75	0.67	3.59	0.70	-0.24	3.68	0.70	3.44	0.69	-0.34
Goal Expectations	3.65	0.57	3.57	0.59	-0.15	3.58	0.61	3.69	0.54	0.18
Narcissism	2.55	0.92	2.83	0.78	0.31	2.78	0.80	2.80	0.71	0.02
Tolerance for Ambiguity	2.90	0.50	2.75	0.48	-0.31	2.80	0.49	2.74	0.49	-0.11
Social Acuity	3.75	0.64	3.79	0.54	0.06	3.75	0.55	3.93	0.47	0.32
Work Values										
Benevolence (Likert)	4.22	0.60	4.23	0.55	0.01	4.24	0.53	4.32	0.55	0.16
Social Work Environ (Likert)	3.76	0.68	3.83	0.60	0.10	3.81	0.61	3.86	0.57	0.07
Selfless Service (Rank)	0.72	1.25	0.80	1.18	0.06	0.81	1.18	0.85	1.05	0.03
Leadership (Rank)	-0.14	1.42	0.23	1.17	0.26	0.13	1.21	0.52	0.98	0.32
Recognition (Rank)	0.04	1.14	-0.01	1.08	-0.04	0.02	1.08	-0.09	1.07	-0.11
Pay (Rank)	-0.02	1.08	0.07	1.03	0.08	0.10	1.06	0.01	0.92	-0.09
Structure (Rank)	0.13	1.14	-0.18	1.06	-0.28	-0.08	1.05	-0.22	0.98	-0.14
Comfort (Rank)	-0.08	1.15	-0.28	1.05	-0.17	-0.21	1.09	-0.35	0.99	-0.12
Home (Rank)	-0.14	0.96	-0.12	0.99	0.02	-0.12	1.02	-0.21	0.82	-0.09
Challenge (Rank)	-0.24	0.92	0.05	0.98	0.32	0.04	0.95	-0.07	0.96	-0.12
Self-Direction (Rank)	-0.14	0.83	-0.11	0.91	0.04	-0.09	0.90	-0.27	1.04	-0.20
Teamwork (Rank)	0.01	0.65	-0.15	0.92	-0.24	-0.11	0.87	-0.22	0.92	-0.12
Variety (Rank)	0.14	0.86	0.11	0.87	-0.03	0.07	0.87	0.17	0.78	0.11

Table B.3. (Continued)

Scale/Predictor	OCS Option					Gender				
	In-Service (I)		Enlistment (E)		I-E	Male (M)		Female (F)		M-F
	<i>n</i> = 31 – 36		<i>n</i> = 302 – 326			<i>n</i> = 344 – 376		<i>n</i> = 46 – 54		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>d</i>
Leader Knowledge Test (LKT)										
Traits	9.49	0.09	9.45	0.16	-0.51	9.45	0.16	9.50	0.10	0.29
Skills	9.42	0.12	9.39	0.14	-0.19	9.36	0.13	9.41	0.12	0.39
Tailored Adaptive Personality Assessment System (TAPAS)										
Even Temper	0.26	0.52	0.12	0.59	-0.28	0.16	0.56	-0.02	0.62	-0.31
Curiosity/Continuous Learning	-0.14	0.71	-0.03	0.70	0.16	-0.01	0.71	-0.11	0.57	-0.15
Tolerance	-0.41	0.83	-0.61	0.86	-0.25	-0.62	0.84	-0.14	0.79	0.58
Trust/Cooperation	-0.57	0.47	-0.56	0.55	0.01	-0.58	0.53	-0.57	0.48	0.03
Optimism	0.24	0.63	0.23	0.70	-0.02	0.22	0.68	0.19	0.69	-0.03
Adjustment	0.16	0.64	-0.05	0.68	-0.34	-0.02	0.66	-0.39	0.67	-0.55
Dominance/Leadership	0.19	0.64	0.07	0.59	-0.20	0.07	0.60	0.24	0.57	0.28
Physical Conditioning	0.37	0.90	0.52	0.76	0.17	0.52	0.76	0.53	0.76	0.01
Achievement	0.25	0.59	0.34	0.62	0.14	0.31	0.64	0.41	0.54	0.16
Non-Delinquency	-0.17	0.62	-0.15	0.59	0.03	-0.17	0.59	-0.06	0.53	0.18
Responsibility	-0.21	0.56	-0.08	0.60	0.23	-0.10	0.60	-0.16	0.55	-0.10
Intellectual Efficiency	0.05	0.63	0.12	0.73	0.11	0.10	0.73	0.20	0.62	0.13
Army Identity Structure										
Overlap	5.65	1.10	5.41	0.11	-0.22	5.42	1.14	5.62	0.97	0.18
Concept	4.59	1.62	4.28	1.48	-0.19	4.32	1.49	4.28	1.36	-0.02
Conflict	5.38	1.30	5.04	1.40	-0.26	5.03	1.39	5.15	1.32	0.09

$d = (M_{\text{COMPARISON}} - M_{\text{REFERENT}})/SD_{\text{POOLED}}$, where the referent group is the majority group (Male, White, Non-Hispanic, and White). Results are based on the raw data combining the Phase 2 and Phase 3 samples. Bolded values indicate statistical significance at $p < .05$. Significance is based on the independent samples t -test difference between the two means.

Table B.3. (Continued)

Scale/Predictor	Ethnicity					Race				
	White, Non-Hispanic (WNH)		Hispanic (H)		WNH-H	White (W)		Black (B)		W-B
	<i>n</i> = 284 - 305		<i>n</i> = 12 - 14			<i>n</i> = 319 - 345		<i>n</i> = 34 - 42		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
AFQT	87.87	9.74	80.64	11.42	-0.74	87.68	9.98	79.19	15.22	-0.85
Rational Biodata Inventory (RBI)										
Peer Leadership	3.98	0.54	3.89	0.73	-0.15	3.97	0.55	4.06	0.49	0.16
Achievement	4.21	0.45	4.13	0.82	-0.18	4.20	0.45	4.25	0.60	0.13
Fitness Motivation	3.96	0.59	3.81	0.36	-0.25	3.95	0.59	3.94	0.63	-0.02
Hostility to Authority	1.87	0.48	2.00	0.50	0.28	1.88	0.48	1.77	0.53	-0.23
Generalized Self-Efficacy	4.32	0.47	4.36	0.44	0.07	4.32	0.46	4.46	0.50	0.32
Affective Commitment	4.03	0.59	4.04	0.56	0.01	4.02	0.59	4.13	0.57	0.18
Stress Tolerance	3.32	0.49	3.20	0.82	-0.26	3.31	0.49	3.32	0.52	0.03
ARC Hostility to Authority	2.83	0.62	2.91	0.96	0.14	2.83	0.61	2.57	0.70	-0.43
Interest in Leadership	4.01	0.69	3.89	0.81	-0.17	3.99	0.69	3.96	0.78	-0.05
Leader Self-Efficacy	4.13	0.44	4.14	0.45	0.03	4.12	0.44	4.20	0.49	0.19
Equity Sensitivity	2.29	0.54	2.27	0.74	-0.04	2.29	0.55	2.17	0.62	-0.21
Tolerance for Injury	3.80	0.58	3.57	0.75	-0.41	3.80	0.57	3.58	0.74	-0.38
Emotional Stability	3.65	0.68	3.53	0.99	-0.18	3.64	0.68	3.76	0.73	0.17
Goal Expectations	3.57	0.60	3.69	0.48	0.20	3.56	0.59	3.72	0.64	0.26
Narcissism	2.75	0.76	2.93	0.89	0.23	2.78	0.75	2.61	0.88	-0.22
Tolerance for Ambiguity	2.79	0.46	2.96	0.75	0.36	2.78	0.46	2.79	0.47	0.01
Social Acuity	3.75	0.54	3.86	0.59	0.19	3.76	0.54	3.88	0.47	0.23
Work Values										
Benevolence (Likert)	4.23	0.52	4.39	0.49	0.29	4.23	0.52	4.24	0.60	0.03
Social Work Environ (Likert)	3.82	0.60	3.91	0.48	0.17	3.82	0.59	3.78	0.75	-0.08
Selfless Service (Rank)	0.79	1.19	1.22	1.01	0.36	0.79	1.17	0.95	1.18	0.14
Leadership (Rank)	0.22	1.16	0.31	1.32	0.08	0.20	1.17	0.33	1.08	0.12
Recognition (Rank)	0.02	1.08	0.32	1.28	0.28	0.00	1.07	-0.05	1.20	-0.05
Pay (Rank)	0.08	1.06	-0.25	0.97	-0.31	0.07	1.04	0.01	0.93	-0.05
Structure (Rank)	-0.07	1.01	0.28	1.12	0.34	-0.08	1.01	-0.27	1.18	-0.19
Comfort (Rank)	-0.23	1.06	0.37	1.20	0.57	-0.28	1.07	-0.02	1.02	0.25
Home (Rank)	-0.06	0.99	-0.34	1.16	-0.28	-0.07	0.97	-0.37	0.95	-0.31
Challenge (Rank)	0.04	0.95	0.31	1.03	0.28	0.01	0.97	0.04	0.76	0.03
Self-Direction (Rank)	-0.10	0.92	-0.20	0.94	-0.11	-0.10	0.90	-0.26	0.94	-0.18
Teamwork (Rank)	-0.13	0.88	0.17	0.66	0.34	-0.13	0.88	-0.12	0.75	0.02
Variety (Rank)	0.07	0.87	0.30	0.60	0.26	0.08	0.87	0.33	0.75	0.29

Table B.3. (Continued)

Table D.3: (Continued)

Scale/Predictor	Ethnicity					Race				
	White, Non-Hispanic (WNH)		Hispanic (H)		WNH-H	White (W)		Black (B)		W-B
	<i>n</i> = 305 - 284		<i>n</i> = 14 - 12			<i>n</i> = 345 - 319		<i>n</i> = 42 - 34		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Leader Knowledge Test (LKT)										
Traits	9.40	0.13	9.36	0.18	-0.23	9.45	0.16	9.48	0.11	0.16
Skills	9.45	0.16	9.40	0.22	-0.29	9.40	0.13	9.39	0.12	-0.02
Tailored Adaptive Personality Assessment System (TAPAS)										
Even Temper	0.15	0.58	-0.14	0.53	-0.50	0.14	0.58	0.04	0.50	-0.18
Curiosity/Continuous Learning	0.00	0.71	-0.18	0.75	-0.25	-0.01	0.71	-0.08	0.64	-0.11
Tolerance	-0.63	0.85	-0.19	0.81	0.51	-0.63	0.86	-0.35	0.72	0.33
Trust/Cooperation	-0.58	0.55	-0.67	0.36	-0.16	-0.58	0.54	-0.59	0.42	-0.01
Optimism	0.19	0.67	0.37	0.71	0.27	0.16	0.67	0.46	0.71	0.44
Adjustment	-0.09	0.66	0.16	0.86	0.39	-0.11	0.64	0.21	0.79	0.49
Dominance/Leadership	0.12	0.58	0.12	0.80	-0.01	0.11	0.59	0.09	0.58	-0.03
Physical Conditioning	0.56	0.76	0.48	0.80	-0.11	0.56	0.77	0.42	0.87	-0.18
Achievement	0.36	0.62	0.33	0.69	-0.05	0.34	0.62	0.30	0.61	-0.07
Non-Delinquency	-0.17	0.59	-0.22	0.58	-0.09	-0.16	0.59	-0.12	0.47	0.07
Responsibility	-0.11	0.59	-0.36	0.60	-0.42	-0.10	0.60	-0.08	0.55	0.04
Intellectual Efficiency	0.12	0.73	0.10	0.68	-0.03	0.14	0.73	0.00	0.73	-0.20
Army Identity Structure										
Overlap	5.43	1.13	5.53	0.99	0.09	5.42	1.13	5.83	1.06	0.36
Concept	4.34	1.49	4.33	1.54	-0.01	4.31	1.47	4.55	1.50	0.17
Conflict	5.07	1.37	4.53	1.51	-0.39	5.06	1.35	5.28	1.38	0.16

Note. *d* = Cohen's *d*. Coefficients in bold are statistically significant ($p < .05$) using an independent samples *t*-test. $d = (M_{\text{COMPARISON}} - M_{\text{REFERENT}})/SD_{\text{POOLED}}$, where the referent group is the majority group (Male, White, Non-Hispanic, and White). Results are based on the raw data combining the Phase 2 and Phase 3 samples. Bolded values indicate statistical significance at $p < .05$. Significance is based on the independent samples *t*-test difference between the two means.

Appendix C

Bivariate Predictor and Criterion Intercorrelations for the SelectOCS Phase 2 and 3 Samples

Table C.1. (Continued)

Scale	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1. RBl: Peer Leadership	.155	.001	-.044	.030	.025	.030	.030	.121	.049	.065	-.088	.030	.150	-.128	.175
2. RBl: Achievement	.145	.067	.009	-.013	.076	-.026	.142	.127	.123	.063	-.016	.130	.107	-.183	.037
3. RBl: Fitness Motivation	.120	.080	-.003	-.026	.013	.038	.014	.028	.063	.115	-.063	.002	.040	-.102	.094
4. RBl: Hostility to Authority	-.042	.075	.044	.051	-.079	-.056	-.126	-.043	-.040	-.009	-.052	.016	-.062	.055	.018
5. RBl: Generalized Self-Efficacy	.039	.008	-.063	-.042	.033	.015	.100	.111	.084	.096	-.056	.012	.008	-.101	.280
6. RBl: Affective Commitment	.047	.055	-.117	.016	.027	.016	.075	.006	.045	.084	-.058	.047	-.052	-.038	.105
7. RBl: Stress Tolerance	-.064	-.181	-.129	-.047	.039	.012	.091	.022	.053	.068	-.021	-.084	.065	-.077	.198
8. RBl: ARC Hostility to Authority	-.036	.138	.025	-.056	-.020	-.024	-.156	-.077	-.047	-.053	-.019	-.020	-.100	-.062	-.125
9. RBl: Interest in Leadership	.144	.079	-.024	.074	.079	.101	.094	.174	.123	.150	-.144	.108	-.008	-.158	.152
10. RBl: Leader Self-Efficacy	.119	.084	-.073	.085	.083	.042	.090	.139	.114	.148	-.080	.051	.024	-.169	.206
11. RBl: Equity Sensitivity	.008	.084	.047	.101	.014	.021	-.052	.014	-.140	-.062	-.078	.052	-.222	-.075	-.152
12. RBl: Tolerance for Injury	-.018	-.034	-.108	-.009	-.052	-.011	.066	.030	.104	.085	-.043	.090	.001	-.174	.052
13. Values (Likert):Benevolence	.202	.034	.016	.026	-.037	-.066	.014	.087	.134	-.006	-.020	.166	.219	-.006	.061
14. Values (Likert):Social Work Environment	.108	.031	-.050	.051	.018	-.015	.041	.091	.186	.066	-.033	.053	.159	.148	.122
15. Values (Rank):Selfless Service	.291	.188	.215	.134	.181	.167	.205	.372	.431	.326	-.019	.068	.183	-.005	.024
16. Values (Rank):Leadership	-	.302	.217	.301	.208	.243	.284	.339	.276	.354	-.206	.020	-.011	-.036	.001
17. Values (Rank):Recognition	.374	-	.260	.202	.217	.209	.233	.250	.255	.288	-.192	-.036	-.053	.027	.035
18. Values (Rank):Pay	.182	.320	-	.340	.187	.246	.182	.241	.271	.308	-.052	-.063	-.060	.028	-.076
19. Values (Rank):Structure	.166	.198	.378	-	.348	.312	.296	.280	.294	.415	-.057	.066	-.135	-.077	-.047
20. Values (Rank):Comfort	.222	.237	.212	.408	-	.361	.238	.273	.309	.318	-.072	.052	-.088	.050	.070
21. Values (Rank):Home	.251	.216	.217	.220	.435	-	.352	.274	.289	.369	-.063	.117	-.147	-.066	-.108
22. Values (Rank):Challenge	.190	.242	.271	.233	.264	.427	-	.483	.415	.422	-.063	.080	.007	-.029	.010
23. Values (Rank):Self-Direction	.258	.253	.277	.243	.252	.305	.472	-	.486	.464	-.061	.116	.022	-.004	-.020
24. Values (Rank):Teamwork	.285	.271	.238	.184	.293	.292	.377	.545	-	.547	-.051	.109	.056	-.035	.058
25. Values (Rank):Variety	.279	.301	.305	.338	.296	.341	.419	.399	.453	-	-.166	.017	-.091	.005	.059
26. TAPAS: Even Temper	-.108	-.053	-.013	-.018	-.048	-.042	-.021	-.087	-.027	-.138	-	-.030	.019	.133	.137
27. TAPAS: Curiosity/Continuous Learning	-.006	-.058	-.014	.000	-.043	.063	.059	.096	.092	.107	-.021	-	.152	-.131	-.061
28. TAPAS: Tolerance	-.023	-.027	-.044	-.011	-.094	-.092	.011	.016	.052	-.023	.068	.269	-	.025	.034
29. TAPAS: Trust/Cooperation	.005	.015	.007	-.032	.014	-.049	.013	-.126	-.092	-.076	.213	-.085	.109	-	.077
30. TAPAS: Optimism	-.007	-.010	-.045	-.022	-.048	-.020	-.026	-.075	-.016	.014	.151	.017	.039	.083	-
31. TAPAS: Adjustment	-.079	-.033	-.046	-.055	.003	.000	-.032	-.040	-.005	.074	.090	-.014	-.143	-.027	.270
32. TAPAS: Dominance/Leadership	.077	-.019	-.099	-.077	.015	-.025	-.059	.033	-.003	-.023	-.278	.099	.068	-.175	.149
33. TAPAS: Physical Conditioning	-.006	-.053	-.048	-.067	-.067	-.063	-.045	-.040	.040	.026	-.125	-.037	-.094	-.154	.043
34. TAPAS: Achievement	.040	.072	-.043	-.055	-.033	.085	.079	.091	.105	.088	-.064	.126	-.057	-.015	-.079
35. TAPAS: Non-delinquency	.056	.052	.037	-.085	-.050	.005	.002	-.096	-.088	-.051	.142	-.070	.075	.230	-.080
36. TAPAS: Responsibility	-.009	-.041	-.059	-.030	.016	-.025	-.039	-.045	.019	.027	.134	-.016	.066	.012	.073
37. TAPAS: Intellectual Efficiency	-.077	-.095	-.106	-.024	-.022	-.042	-.008	.020	-.058	-.059	-.032	.429	.171	-.028	.073
38. Army Identity Structure: Overlap	.004	.032	-.039	.005	.034	.055	.071	.076	.130	.058	-.048	.085	-.094	-.069	.065
39. Army Identity Structure: Concept	.077	.108	-.003	.001	-.007	.024	.096	.035	.110	.146	-.055	.050	-.098	.009	.000
40. Army Identity Structure: Conflict	.057	.056	-.016	-.049	-.060	.044	.061	.076	.089	.022	-.047	.030	-.045	.026	.081
41. Career Intentions	.053	-.015	.003	-.017	.045	-.013	.080	.042	.049	-.008	-.029	-.052	.049	.031	.034
42. Affective Commitment	.027	.020	.013	.008	.059	.086	.163	.085	.136	.088	.050	-.018	.015	.162	.109
43. Peer Ratings of Leadership Potential	-.037	-.038	-.040	-.021	.037	.051	.028	.003	.044	.094	-.042	-.112	-.073	-.091	.112
44. OCS Total Score	.021	.010	-.041	-.018	-.004	.089	.033	.014	.048	.148	-.105	-.081	-.050	-.075	.025
45. OCS Leadership Score	-.003	.000	-.036	.022	.050	.091	.031	-.027	.024	.167	-.076	-.130	-.072	-.034	.053
46. OCS Physical Fitness	.080	.053	-.045	-.058	-.066	.065	.019	.069	.083	.132	-.103	-.020	-.053	-.141	.005
47. OCS Academic Performance	-.031	-.013	.024	-.052	.000	.030	.021	.013	.019	-.038	-.003	.103	.079	.005	-.056
48. OCS Average History Score	-.066	-.043	.021	-.025	.034	.063	.003	.023	.042	-.084	-.029	.090	.079	-.020	-.077
49. OCS APFT Score	.078	.045	-.025	-.048	-.084	.045	.023	.101	.067	.100	-.091	-.025	-.029	-.143	-.040
50. AFQT	-.071	-.067	-.036	-.030	.001	-.002	-.001	.028	-.047	-.040	.050	.196	.142	.121	-.011
51. LKT: Traits	-.065	.010	-.094	-.070	-.068	-.037	.141	-.041	-.019	.000	-.049	.070	.042	.099	-.033
52. LKT: Skills	-.041	.025	-.042	.012	-.012	-.021	.126	-.015	-.022	.008	.009	.122	.100	.049	-.014

Table C.1. (Continued)

Scale		46	47	48	49	50	51	52
1	RBI: Peer Leadership	.152	-.042	-.043	.165	-.003	.013	.061
2	RBI: Achievement	.071	-.070	-.071	.101	.031	.035	.054
3	RBI: Fitness Motivation	.566	.105	.061	.528	-.060	-.017	-.013
4	RBI: Hostility to Authority	-.175	-.076	.027	-.206	-.064	-.136	-.122
5	RBI: Generalized Self-Efficacy	.152	-.111	-.107	.178	-.025	-.005	.031
6	RBI: Affective Commitment	.035	-.026	.048	.070	-.087	.062	.022
7	RBI: Stress Tolerance	.107	.183	.125	.080	.010	.079	.109
8	RBI: ARC Hostility to Authority	-.069	-.119	-.120	-.049	-.052	-.150	-.137
9	RBI: Interest in Leadership	.154	-.011	.000	.158	-.008	.039	.039
10	RBI: Leader Self-Efficacy	.175	-.073	-.031	.184	-.037	.023	.044
11	RBI: Equity Sensitivity	-.120	-.131	-.075	-.111	-.063	-.140	-.072
12	RBI: Tolerance for Injury	.308	.214	.202	.267	-.020	.002	-.009
13	Values (Likert):Benevolence	.057	-.023	-.003	.095	-.036	-.020	-.020
14	Values (Likert):Social Work Environment	-.011	-.092	-.039	-.017	-.149	-.015	-.005
15	Values (Rank):Selfless Service	.001	.025	-.004	.031	-.033	.043	-.016
16	Values (Rank):Leadership	-.023	-.082	-.119	.019	-.020	-.021	-.012
17	Values (Rank):Recognition	.031	-.046	-.004	.055	.062	.017	.012
18	Values (Rank):Pay	-.047	-.051	-.032	-.037	-.008	.016	-.076
19	Values (Rank):Structure	-.013	.105	.042	-.064	.047	.048	.052
20	Values (Rank):Comfort	-.001	-.011	-.039	.019	.067	-.034	-.036
21	Values (Rank):Home	.084	.033	.005	.042	.063	.028	-.024
22	Values (Rank):Challenge	-.047	.089	.021	-.050	.041	.014	.008
23	Values (Rank):Self-Direction	-.037	.007	-.034	-.032	-.040	-.103	-.029
24	Values (Rank):Teamwork	.065	.174	.076	.040	.049	-.054	-.066
25	Values (Rank):Variety	.089	.154	.076	.066	.044	.060	.081
26	TAPAS: Even Temper	-.026	.034	.062	-.109	.019	.021	-.037
27	TAPAS: Curiosity/Continuous Learning	-.103	.056	.104	-.080	.202	.043	.003
28	TAPAS: Tolerance	.076	.065	.102	.074	.042	.080	.072
29	TAPAS: Trust/Cooperation	-.122	-.027	-.024	-.104	-.063	.085	.032
30	TAPAS: Optimism	.003	.057	.055	-.015	-.051	.122	.137
31	TAPAS: Adjustment	.012	.151	.163	-.025	.046	-.031	.022
32	TAPAS: Dominance/Leadership	.024	.058	.036	.073	.070	.051	.089
33	TAPAS: Physical Conditioning	.386	-.050	-.082	.373	-.113	-.030	-.046
34	TAPAS: Achievement	.139	.098	-.036	.139	-.009	.118	.085
35	TAPAS: Non-delinquency	-.041	-.037	-.063	-.008	-.036	.100	.054
36	TAPAS: Responsibility	.044	.069	.049	.070	.006	.122	.117
37	TAPAS: Intellectual Efficiency	-.131	.140	.147	-.130	.277	.056	.132
38	Army Identity Structure: Overlap	-.040	-.109	-.059	-.004	-.102	.121	.099
39	Army Identity Structure: Concept	.008	-.019	-.043	.016	-.108	.066	.025
40	Army Identity Structure: Conflict	.035	-.022	-.024	.064	-.102	.096	.072
41	Career Intentions	-.087	-.143	-.185	-.074	-.148	.108	.013
42	Affective Commitment	.000	.012	.028	.041	-.020	.105	.014
43	Peer Ratings of Leadership Potential	.219	.248	.105	.132	.149	.037	.042
44	OCS Total Score	.683	.543	.342	.602	.223	.072	.024
45	OCS Leadership Score	.460	.340	.195	.407	.178	.091	.012
46	OCS Physical Fitness	-	.236	.114	.922	-.031	-.009	.025
47	OCS Academic Performance	.199	-	.758	.148	.461	.036	.035
48	OCS Average History Score	.141	.824	-	.061	.361	-.009	-.027
49	OCS APFT Score	.958	.204	.144	-	-.018	-.026	.010
50	AFQT	.068	.509	.506	.070	-	.138	.146
51	LKT: Traits	.041	.099	.065	.035	.165	-	.689
52	LKT: Skills	.011	.170	.124	-.014	.244	.660	-

Note. $N = 429$. Correlations for Phase 2 sample appear below the diagonal; correlations for Phase 3 sample appear above the diagonal. Correlations are corrected using Full Information Maximum Likelihood (FIML) missing data estimation. Correlations greater than or equal to .095 are statistically significant.

Appendix D

Full Information Maximum Likelihood (FIML) Bivariate Correlations for Phase 3 In-Service and Enlistment Option Candidates

Table D.1. (Continued)

Scale	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1 AFQT	--	--	--	--	--	-.093	-.138	-.220	-.115	-.127	-.072	-.052	-.129	.016	-.068
2 RBI: Peer Leadership	--	--	--	--	--	.169	-.025	-.008	-.095	-.091	-.094	.031	-.088	.138	.179
3 RBI: Achievement	--	--	--	--	--	.105	-.038	.083	-.021	-.046	-.033	.116	.088	.107	.138
4 RBI: Fitness Motivation	--	--	--	--	--	.032	-.076	-.034	-.082	-.001	-.099	.046	.116	.026	.037
5 RBI: Hostility to Authority	--	--	--	--	--	-.014	.076	.024	.064	.169	.024	.000	-.051	.024	-.094
6 RBI: Generalized Self-Efficacy	--	--	--	--	--	.124	-.049	-.056	-.095	-.119	-.110	-.041	-.013	.065	.155
7 RBI: Affective Commitment	--	--	--	--	--	.148	.049	.131	-.013	-.079	.041	.086	.120	.087	.211
8 RBI: Stress Tolerance	--	--	--	--	--	-.083	-.082	-.099	-.163	-.107	-.052	-.039	-.030	.016	-.041
9 RBI: ARC Hostility to Authority	--	--	--	--	--	-.034	.001	.019	.110	.081	.034	.103	.043	-.025	-.006
10 RBI: Interest in Leadership	--	--	--	--	--	.112	.086	-.019	-.119	-.085	-.043	.074	.010	.107	.020
11 RBI: Leader Self-Efficacy	--	--	--	--	--	.136	-.014	-.006	-.092	-.145	-.089	.038	.011	.044	.137
12 RBI: Equity Sensitivity	--	--	--	--	--	-.097	-.055	-.021	.003	.041	.010	-.023	.022	-.017	-.180
13 RBI: Tolerance for Injury	--	--	--	--	--	-.028	-.036	-.174	-.206	-.255	-.115	.010	-.015	.088	.032
14 Values (Likert): Benevolence	--	--	--	--	--	.391	.132	.091	.089	-.028	-.010	.012	.034	.102	.181
15 Values (Likert): Social Work Environment	--	--	--	--	--	.228	.043	.165	.129	-.033	-.006	.131	.152	.238	.098
16 RBI: Emotional Stability	-	--	--	--	--	--	--	--	--	--	--	--	--	--	--
17 RBI: Goal Expectations	.205	-	--	--	--	--	--	--	--	--	--	--	--	--	--
18 RBI: Narcissism	-.305	.150	--	--	--	--	--	--	--	--	--	--	--	--	--
19 RBI: Tolerance for Ambiguity	.501	.012	-.317	-	--	--	--	--	--	--	--	--	--	--	--
20 RBI: Social Acuity	.202	.290	.048	.223	-	--	--	--	--	--	--	--	--	--	--
21 Values (Rank):Selfless Service	-.013	.035	.127	.018	.096	-	.419	.336	.209	.152	.165	.256	.219	.283	.320
22 Values (Rank):Leadership	-.018	.141	.147	-.070	.096	.328	-	.461	.297	.285	.241	.250	.280	.350	.405
23 Values (Rank):Recognition	-.019	.050	.176	-.113	.065	.234	.344	-	.435	.306	.319	.313	.210	.215	.341
24 Values (Rank):Pay	-.086	.004	.086	-.137	-.044	.169	.198	.294	-	.386	.258	.248	.129	.236	.313
25 Values (Rank):Structure	.063	-.003	.018	.029	-.050	.146	.226	.201	.364	-	.370	.256	.240	.312	.322
26 Values (Rank):Comfort	.104	.038	.109	.000	-.047	.202	.216	.228	.203	.386	-	.457	.329	.237	.320
27 Values (Rank):Home	.123	.114	.052	.027	-.077	.210	.249	.216	.231	.256	.406	-	.406	.310	.364
28 Values (Rank):Challenge	.144	.067	-.036	.021	-.015	.231	.233	.241	.235	.260	.254	.397	-	.500	.380
29 Values (Rank):Self-Direction	.090	.102	.031	-.037	-.017	.328	.295	.253	.264	.259	.261	.293	.477	-	.440
30 Values (Rank):Teamwork	.149	.073	-.051	.086	.058	.376	.286	.266	.256	.237	.305	.289	.394	.520	-
31 Values (Rank):Variety	.151	.098	-.091	.048	-.020	.358	.312	.298	.306	.370	.304	.351	.422	.426	.495
32 LKT: Traits	.081	.004	-.108	.064	.066	.005	-.047	.012	-.047	-.014	-.052	-.015	.075	-.070	-.034
33 LKT: Skills	.110	.081	-.116	.055	.119	-.007	-.032	.020	-.055	.033	-.023	-.022	.064	-.026	-.042
34 TAPAS: Even Temper	.065	-.129	-.037	.023	-.081	-.029	-.153	-.114	-.029	-.032	-.054	-.050	-.037	-.077	-.035
35 TAPAS: Curiosity/Continuous Learning	-.006	.077	.024	-.028	-.030	.042	.011	-.048	-.030	.032	.000	.080	.066	.108	.101
36 TAPAS: Tolerance	-.027	-.016	-.095	.014	.107	.077	-.013	-.034	-.049	-.062	-.087	-.115	.007	.013	.054
37 TAPAS: Trust/Cooperation	-.107	-.175	-.041	.017	-.068	.004	-.010	.018	.008	-.045	.029	-.054	-.004	-.079	-.065
38 TAPAS: Optimism	.369	.128	.002	.191	.221	-.033	-.017	.007	-.062	-.038	-.010	-.054	-.008	-.057	.017
39 TAPAS: Adjustment	.335	.032	-.132	.142	.005	-.072	-.105	-.049	-.097	-.072	-.016	-.046	-.032	-.048	.001
40 TAPAS: Dominance/Leadership	.094	.298	.094	.067	.361	.036	.108	.010	-.103	-.027	.003	-.023	-.028	.055	.036
41 TAPAS: Physical Conditioning	.102	.174	.035	.066	.048	-.049	.032	-.020	-.034	-.048	-.053	-.041	-.026	-.003	.026
42 TAPAS: Achievement	.121	.187	-.077	.023	.009	.028	.040	.013	-.018	-.029	-.011	.050	.065	.085	.094
43 TAPAS: Non-delinquency	-.132	.019	.079	-.174	-.073	.081	.006	.021	.019	-.096	-.061	-.003	-.021	-.069	-.069
44 TAPAS: Responsibility	.114	.052	-.069	.107	.212	.048	-.016	-.042	-.049	-.045	.009	-.052	-.023	.000	.022
45 TAPAS: Intellectual Efficiency	.098	.057	-.064	.166	.118	-.083	-.088	-.097	-.121	.003	-.011	-.050	.015	.027	.007
46 Army Identity Structure: Overlap	.188	.321	.018	.110	.105	.118	.035	.042	-.018	.032	.018	.083	.105	.078	.109
47 Army Identity Structure: Concept	.132	.172	.021	.080	.096	.114	.024	.068	-.021	-.016	-.056	.008	.064	.043	.059
48 Army Identity Structure: Conflict	.227	.252	-.058	.158	.107	.112	.057	-.002	-.036	-.059	-.087	.036	.058	.054	.041
49 Career Intentions	.105	.260	-.037	.072	.109	.067	.045	.004	.034	-.003	.039	.036	.087	.057	.058
50 Affective Commitment	.077	.130	-.021	.180	.142	.174	.052	.040	.001	.021	.052	.087	.128	.050	.096
51 Peer Ratings of Leadership Potential	.195	.161	-.024	.017	.121	.054	.002	-.032	-.041	-.015	.026	.049	.033	.026	.092
52 OCS History Course Score	-.001	.014	.032	.097	.007	.026	-.071	-.025	-.001	.001	.000	.027	.011	.011	.059
53 OCS APFT Score	.104	.174	-.023	.040	.054	.055	.064	.043	-.044	-.062	-.048	.059	.008	.056	.047

Table D.1. (Continued)

Scale	46	47	48	49	50	51	52	53
1 AFQT	-.273	-.147	-.250	-.085	.093	.212	.603	-.201
2 RBI: Peer Leadership	.103	.120	.224	.056	.210	.058	.122	-.073
3 RBI: Achievement	.237	.180	.306	.154	.187	.077	-.001	.107
4 RBI: Fitness Motivation	.112	.123	.149	.087	.126	.166	.163	.208
5 RBI: Hostility to Authority	-.134	.010	-.249	-.228	-.147	-.148	.078	-.059
6 RBI: Generalized Self-Efficacy	.202	.198	.317	.235	.199	.080	.029	-.058
7 RBI: Affective Commitment	.557	.468	.511	.169	.444	.179	.039	-.020
8 RBI: Stress Tolerance	.124	.078	.299	.170	.207	.207	.209	-.156
9 RBI: ARC Hostility to Authority	-.175	-.200	-.312	-.207	-.201	-.180	-.050	-.017
10 RBI: Interest in Leadership	.203	.219	.287	.212	.073	.047	.047	-.057
11 RBI: Leader Self-Efficacy	.221	.204	.363	.167	.154	.136	.016	-.011
12 RBI: Equity Sensitivity	-.212	-.129	-.310	-.263	-.304	-.236	-.024	.036
13 RBI: Tolerance for Injury	.117	.209	.205	.102	.177	.286	.203	.009
14 Values (Likert): Benevolence	.316	.243	.327	.103	.279	.144	-.128	.038
15 Values (Likert): Social Work Environment	.178	.216	.235	.053	.139	.019	-.005	.001
16 RBI: Emotional Stability	--	--	--	--	--	--	--	--
17 RBI: Goal Expectations	--	--	--	--	--	--	--	--
18 RBI: Narcissism	--	--	--	--	--	--	--	--
19 RBI: Tolerance for Ambiguity	--	--	--	--	--	--	--	--
20 RBI: Social Acuity	--	--	--	--	--	--	--	--
21 Values (Rank):Selfless Service	.123	.060	.232	.033	.129	-.040	-.132	-.034
22 Values (Rank):Leadership	.127	.081	.185	.077	-.043	-.074	-.116	-.106
23 Values (Rank):Recognition	.130	-.023	.102	.039	.007	-.024	-.201	.085
24 Values (Rank):Pay	-.009	-.035	.027	-.197	.041	-.053	-.166	.034
25 Values (Rank):Structure	-.017	-.023	-.009	-.047	-.114	-.151	-.198	.060
26 Values (Rank):Comfort	.118	.107	.064	-.082	.058	.008	-.130	-.003
27 Values (Rank):Home	.065	.066	.086	-.038	-.014	.043	-.116	-.059
28 Values (Rank):Challenge	.129	.175	.105	.054	-.021	.034	-.069	.043
29 Values (Rank):Self-Direction	.038	.116	.095	.047	-.060	.048	.074	-.106
30 Values (Rank):Teamwork	.138	.023	.153	.122	.134	-.042	-.121	-.126
31 Values (Rank):Variety	.148	-.012	.118	.195	.129	-.087	-.161	-.109
32 LKT: Traits	-.037	-.081	.063	.108	.166	.041	.135	-.083
33 LKT: Skills	-.141	-.105	.028	.055	-.074	.095	.211	-.047
34 TAPAS: Even Temper	.025	-.093	-.013	.105	.072	.083	-.057	-.032
35 TAPAS: Curiosity/Continuous Learning	.035	.044	-.112	.053	.073	-.102	.153	-.003
36 TAPAS: Tolerance	-.028	-.008	.052	.009	.064	-.098	-.009	-.013
37 TAPAS: Trust/Cooperation	-.025	-.078	-.042	.090	-.008	-.116	-.177	.032
38 TAPAS: Optimism	.149	.166	.115	.056	.079	-.001	-.042	.000
39 TAPAS: Adjustment	-.008	-.041	.103	-.112	.015	-.082	.009	-.030
40 TAPAS: Dominance/Leadership	.181	.248	.282	.156	.056	.075	.015	.013
41 TAPAS: Physical Conditioning	.082	.198	.089	.071	.045	.099	.083	.341
42 TAPAS: Achievement	.160	.177	.213	.145	.069	-.016	.062	.015
43 TAPAS: Non-delinquency	.113	.031	.124	.079	-.017	-.141	-.143	-.026
44 TAPAS: Responsibility	.192	.237	.355	.208	.097	.191	.044	.046
45 TAPAS: Intellectual Efficiency	-.042	.067	.081	.050	.016	-.022	.268	-.128
46 Army Identity Structure: Overlap	-	.620	.509	.177	.307	.084	-.132	-.022
47 Army Identity Structure: Concept	.572	-	.522	.101	.182	.154	-.010	.053
48 Army Identity Structure: Conflict	.564	.500	-	.287	.243	.044	-.099	-.117
49 Career Intentions	.343	.278	.275	-	.207	.168	.069	-.094
50 Affective Commitment	.325	.258	.344	.287	-	.079	.036	-.093
51 Peer Ratings of Leadership Potential	.055	.002	.010	.032	.050	-	.369	.114
52 OCS History Course Score	-.082	-.139	-.105	-.110	-.035	.204	-	-.094
53 OCS APFT Score	.048	.044	.041	-.038	.065	.305	.070	-

Note. $n = 748$ (enlistment option); $n = 241$ (in-service). Correlations are corrected using Full Information Maximum Likelihood (FIML) missing data estimation. Correlations for the enlistment option sample appear below the diagonal; correlations for the in-service sample appear above the diagonal. For the enlistment option, correlations greater than or equal to .071 are statistically significant. For in-service, correlations greater than or equal to .126 are statistically significant. Five scales were added to the RBI for Phase 3 (Emotional Stability, Goal Expectations, Narcissism, Tolerance for Ambiguity, and Social Acuity). These scales were not included in the in-service analyses for Phase 3 due to an insufficient sample size for missing data estimation.